

INSTITUT FÜR DEUTSCHE SPRACHE
FORSCHUNGSBERICHTE

herausgegeben von

Ulrich Engel und Irmgard Vogel

Band 13

Mannheim 1973

© Verlag Gunter Narr · Tübingen 1973

» Tübinger Beiträge zur Linguistik «

74 Tübingen 1 · Postfach 2567

Printed in Germany

Druck: FOTODRUCK PRÄZIS Barbara v. Spangenberg KG

PANTELIS NIKITOPOULOS

STATISTIK FÜR LINGUISTEN

Eine methodische Darstellung

I. Teil

Vorbemerkungen der Herausgeber

Der vorliegende 1. Teil der "Statistik für Linguisten" führt in einen Bereich ein, dessen Relevanz für die Linguistik seit geraumer Zeit unbestritten ist und der in den Forschungen des IdS seit seinem Bestehen eine maßgebliche Rolle gespielt hat. Zugänge zur Statistik gibt es viele. Das IdS sieht es als seine besondere Aufgabe an, diesen Zugang auch dem Linguisten zu vermitteln.

Der Verfasser des Bandes, Pantelis Nikitopoulos, ist seit 1969 wissenschaftlicher Mitarbeiter des IdS. Er hat in Heidelberg Volkswirtschaft studiert und sich jahrelang eingehend mit dem gesamten Bereich der Statistik beschäftigt. Längere Zeit war Pantelis Nikitopoulos in der Abteilung Linguistische Datenverarbeitung des IdS tätig.

Für Kritik und Anregungen sind Autor und Herausgeber dankbar.

Ulrich Engel
Irmgard Vogel

*“Man muß nur in dieser Welt leben,
um zu spüren, was not tut: Gedanken,
Fakten, Gedanken, Zahlen, Gedanken.”*

C. Wright Mills

Kritik der soziologischen Denkweise

“Die kleinen, unerklärten Fakten enthalten in sich immer genug, um alle Erklärungen der großen Fakten zu entkräften.”

Paul Valéry

Windstriche

INHALT

A. Vorwort

B. Deskriptive Statistik

1. Statistische Grundbegriffe

2. Empirische Verteilungen

- 2.1 Organisation statistischer Daten.
Häufigkeitsverteilungen.
- 2.2 Graphische Darstellung.
- 2.3 Verteilung relativer Häufigkeiten.
- 2.4 Kumulative Häufigkeitsverteilung.
- 2.5 Qualitative Merkmale.
- 2.6 Die Lorenz-Kurve.

3. Maßzahlen der Häufigkeitsverteilung

- 3.1 Maßzahlen der zentralen Tendenz.
 - 3.1.1 Das arithmetische Mittel
 - 3.1.2 Der Median (Zentralwert)
 - 3.1.3 Der Modus (Häufigster Wert)
 - 3.1.4 Charakteristische Eigenschaften der
wichtigsten Mittelwerte
- 3.2 Maßzahlen der Streuung
 - 3.2.1 Die Spannweite
 - 3.2.2 Die mittlere lineare Abweichung

- 3.2.3 Die mittlere quadratische Abweichung
(Standardabweichung)
- 3.2.4 Der Variationskoeffizient
- 3.2.5 Der Yule-Herdan-Index

3.3 Maßzahlen der Schiefe

3.4 Maßzahlen der Wölbung

C. Messung und Skalierung

1. Allgemeine Charakterisierungen

2. Skalenarten

- 2.1 Die Nominalskala
- 2.2 Die Ordinalskala
- 2.3 Die Intervallskala
- 2.4 Die Ratio- oder Verhältnisskala

3. Meßbarkeitsniveau und statistische Operationen

4. Meßbarkeit und wissenschaftliche Argumentation

D. Wahrscheinlichkeitsrechnung

1. Kausalität und Zufall

2. Ereignisraum und Elementarereignis

3. Begriff der Wahrscheinlichkeit

4. Die Axiomatisierung der Wahrscheinlichkeitsrechnung

5. Zusammengesetzte Ereignisse

6. Bedingte Wahrscheinlichkeit und Unabhängigkeit

7. Kombinatorik

8. Nicht-klassische Statistik

- 8.1 Das Boltzmann-Maxwell Modell
- 8.2 Das Bose-Einstein-Modell
- 8.3 Das Fermi-Dirac-Modell
- 8.4 Zusammenfassende Interpretation

9. Informationstheorie

- 9.1 Der informationstheoretische Ansatz
- 9.2 Die statistische Konzeption der Informations-
quantität
- 9.3 Relative Entropie und Redundanz
- 9.4 Informationsgehalt in Sequenzen abhängiger Zeichen
 - 9.4.1 Markoff-Ketten
 - 9.4.2 Bedingte Entropie

A. Vorwort

In der Diskussion über die methodische Ausrichtung der linguistischen Forschung, die durch die umwälzenden Entwicklungen der letzten zwanzig Jahre hervorgerufen wurde, spielt die Frage nach der Brauchbarkeit und Adäquatheit der statistischen Methoden bei der Untersuchung des wie auch immer abgegrenzten linguistischen Objektbereichs eine große Rolle.

Die Kontroverse, die daraus entstanden ist, ist genau so unfruchtbar wie entsprechende in anderen gesellschaftswissenschaftlichen Disziplinen, denn sie wird in der Form von Glaubenskämpfen geführt, ohne eine differenzierte Argumentation auf der Grundlage einer fundierten Kenntnis der erkenntnistheoretischen Problematik, die sowohl die methodische Diversifikation der statistischen Theorie als auch die Tragweite des erreichten Standes der linguistischen Forschung zu berücksichtigen hat.

Die Etikettierung "quantitative" Linguistik für jene Bereiche der sprachwissenschaftlichen Forschung, die mathematische oder statistische Methoden zur Anwendung bringen, geht solange an dem eigentlichen Gegenstand vorbei, solange man nicht bereit ist, den genauen inhaltlichen Hintergrund des "Quantitativen" zu explizieren, d.h. sich das relationale Gefüge und dessen verschiedene Ausprägungen vor Augen zu führen, die die "Quantifizierbarkeit" eines Objektbereichs bedingen. Denn erst hier, in der Explikation der genauen-relationenlogischen Bedingungen der Quantifizierbarkeit, läßt sich über Möglichkeiten einer Erfüllung dieser Bedingungen in konkreten linguistischen Gegenstandsbereichen diskutieren.¹⁾²⁾

Auf der anderen Seite wird die Rolle der Statistik sehr oft mißverstanden. Früher war die Meinung weit verbreitet, daß Statistik die Wissenschaft sei, die sich mit dem Studium von Massenerscheinungen beschäftigt, und diese Auffassung findet man heute noch in zahlreichen Einführungsbüchern. Der entscheidende Unterschied in der modernen Auffassung über die Statistik liegt darin, daß man heute nicht mehr unmittelbar in den Beobachtungen selbst das wesentliche Moment der

Statistik erblickt, sondern in der aktiveren Rolle einer Wissenschaft des Entscheidens angesichts von Ungewißheit d.h. auf Grund unvollständiger Information.³⁾ Das bedeutet natürlich, daß die Anwendung statistischer Methoden sich nicht einfach in datensammelnden Verfahren erschöpfen kann, sondern darüberhinaus vorgängig die objektive Identifikation der jeweiligen Bedingungskonstellationen in den realen Untersuchungsbereichen vornehmen muß. Erst dadurch wird eine explizite Argumentationsweise, die eine Entscheidungsprozedur voraussetzt, überprüfbar angesichts der Ungewißheit, die über die realen Phänomene, d.h. aber der unvollständigen Information (des unvollständigen Abbildes der Wirklichkeit), die dieser Entscheidungsprozedur zu Grunde liegt. So schrieb der Soziologe C. Wright Mills: "Heute würde ich nichts mehr zählen oder messen, bevor ich nicht jedes Element, jede Bedingung und Konsequenz in die Welt beliebiger, von mir kontrollierter Maßstäbe übertragen habe. Das sollte der tiefere Sinn . . . dessen sein, was die Statistiker mit der schrecklichen Phrase 'erst die Welt kennen und dann mit den Erhebungen beginnen' meinen sollten . . .".⁴⁾

Daraus ergibt sich eine aufgeklärte Verbindung zwischen linguistischer Theorie einerseits und statistischer Methodenwissenschaft andererseits, wobei auf die einzelnen epistemologischen Fragen hier nicht eingegangen wird. Zur Entstehung eines neuen Zweigs linguistischer Forschung - aus der erwähnten Verbindung -, und zwar in Anlehnung an andere wissenschaftliche Disziplinen, wie z.B. Ökonometrie, Soziometrie, Biometrie usw. besteht m.E. keine Notwendigkeit.

Die Entstehung einer Glottometrie (oder mit einem anderen Terminus Linguametrie) ist zum jetzigen Zeitpunkt nicht mehr erforderlich, denn der programmatische Anstoß, der bei der Konstitutierung der beispielhaften Disziplinen im Vordergrund gestanden hat, ist in der Linguistik schon vorweggenommen worden. Die Nachteile der Separierung eines neuen Zweigs würde auf der anderen Seite eine weitere Lockerung der Interdependenz der Elemente und Teilbereiche des komplexen linguistischen Gegenstands in der Forschung mit sich bringen, als es manchmal ohnehin der Fall ist.

Die Anwendung statistischer Methoden in der linguistischen Forschung

birgt keine Geheimnisse in sich, sondern nur die genaue Kenntnis der Strukturierung des linguistischen Objektbereichs durch die Theorie und die genaue Kenntnis der statistischen Methoden und ihrer Implikationen. Nur in diesem Fall kann man über die Brauchbarkeit und Adäquatheit statistischer Methoden für linguistische Objektbereiche unter der jeweiligen Reichweite der linguistischen Theorie im jeweiligen *hic et nunc* entscheiden. ⁵⁾

Im vorliegenden I. Teil dieser Arbeit werden in einem I. Kapitel die Organisierung, die Darstellung und die parametrische Beschreibung des statistischen Materials in einiger Ausführlichkeit behandelt. Dieser Problemkreis wird in den einschlägigen Büchern über Statistik für Linguisten meistens sehr fragmentarisch behandelt. Für die praktische Arbeit ist es andererseits unerlässlich, diese methodischen Fragen zu beherrschen.

In einem II. Kapitel werden die sehr wichtigen Fragen der Messung und Skalierung, d.h. die verschiedenen Quantifizierungsniveaus und ihre Eigenschaften dargestellt. Dieser Abschnitt bildet m.E. die Grundlage für die Anwendung der verschiedenen statistischen Methoden. Denn erst durch die Überprüfung der theoretischen Strukturierung des linguistischen Objektbereichs im Hinblick darauf, ob die Bedingungen des jeweiligen Quantifizierungsniveaus erfüllt werden, läßt sich über die Anwendbarkeit einzelner statistischer Verfahren konkret etwas sagen.

In einem III. Kapitel werden dann die Grundlagen der Wahrscheinlichkeitstheorie dargestellt, und zwar in einiger Ausführlichkeit, damit man entsprechende Anwendungen in der Literatur verstehen kann. Generell läßt sich sagen, daß die Arbeit darauf ausgerichtet ist, den Linguisten mit statistischen Methoden und Instrumenten vertraut zu machen, die für ihn relevant sind. Das bedeutet natürlich daß aus dem ganzen statistischen Lehrgebäude diejenigen Teile ausgewählt worden sind, die nach Meinung des Verfassers sowohl notwendig sind, im Hinblick auf den momentanen Stand der linguistischen Forschung als auch brauchbar sein werden, wenn man die Tendenz dieser Forschung antizipiert. Insofern ist natürlich ein subjektives

Element in der Auswahlprozedur enthalten. Darüberhinaus ist aber darauf Wert gelegt worden, so viele Methoden und Verfahren darzubieten, daß sich Linguisten Ergebnisse von Nachbardisziplinen wie z.B. Psychologie, Soziologie usw. aneignen können, soweit sie sich in statistischen Untersuchungen niederschlagen.

Daraus ist ersichtlich, daß es sich bei dieser Arbeit nicht um statistische Linguistik handelt und handeln konnte, sondern um Statistik für Linguisten.

Die Aufspaltung dieser Arbeit in Teile hat nicht nur editionstechnische, sondern auch und primär arbeitstechnische Gründe.

B. Deskriptive Statistik

1. Statistische Grundbegriffe

Das ursprüngliche Material jeder statistischen Untersuchung ist die empirische Betrachtung von Massenerscheinungen. Der originäre Informationsträger der Statistik ist die *statistische Einheit* (manchmal auch statistisches Element genannt). Sie bildet den unmittelbaren Gegenstand der statistischen Erhebung.

Eine statistische Einheit tritt in verschiedenen Formen in Erscheinung, die je nach dem Ziel der Untersuchung zugrundegelegt werden. Eine Aussage über eine statistische Einheit erfolgt daher mit Hilfe ihrer *Merkmale* und deren *Modalitäten*. Die Breite der Anwendungsmöglichkeiten der Statistik zeigt sich auch dadurch, daß man als statistische Einheit prinzipiell jeden materiellen oder immateriellen Gegenstand auffassen kann, z.B. Personen, materielle Dinge, sprachliche Gebilde, Handlungen, Verhaltensverläufe, soziale Gebilde und vieles mehr.

Die statistische Einheit "Person" hat beispielsweise als ein mögliches Merkmal das Geschlecht mit den Merkmalsausprägungen (Modalitäten) männlich und weiblich; sie tritt also unter diesem Aspekt (durch das Merkmal charakterisiert) in zwei Formen in Erscheinung. Da die statistische Einheit grundsätzlich mehrere Erscheinungsformen annehmen kann, wird sie auch als (statistische) Veränderliche oder Variable bezeichnet.

Die eindeutige Festlegung einer statistischen Einheit erfolgt durch ihre Definition, die aus drei Identifikationsmerkmalen bestehen muß; nämlich aus

1. einem sachlichen Identifikationsmerkmal (I_S)
2. einem räumlichen " (I_R) und
3. einem zeitlichen " (I_Z)

Wenn wir mit E eine Menge von Einheiten bezeichnen, dann ist

$$E = \{ e \mid e \text{ tr\"agt } I_S, I_R \text{ und } I_Z \} .$$

Wenn wir weiterhin die Menge der Merkmalsmodalitäten von $E \{m_1, m_2, \dots, m_n\}$ mit M bezeichnen, dann ist eine statistische Erhebung eine eindeutige Abbildung von E in M (was man üblicherweise als Funktion bezeichnet), d.h. jedem Element $e \in E$ wird ein und nur ein Element $m \in M$ zugeordnet:

$e \rightarrow m=f(e)$ (lies: e wird durch f das Element
 m zugeordnet)

oder $f : E \rightarrow M$ (lies: durch f wird E in M abgebildet)
oder üblicher

$$f(e) = m$$

Die statistische Variable durchläuft den Argumentenbereich E der Funktion

Die Identifikationsmerkmale sind nicht Gegenstand der statistischen Erhebung; sie haben vielmehr die Aufgabe, die Fragestellung der Erhebung unter formalen Kriterien zu konkretisieren und abzugrenzen, sie bilden infolgedessen die unerläßliche Prämisse der Erhebung im Gegensatz zu den oben besprochenen Merkmalen der statistischen Einheiten (Prädikatsmerkmale), die erhoben werden.

Aus den statistischen Einheiten entsteht die *statistische Masse*, (auch statistische Menge, Grundgesamtheit oder Population genannt). Sie ist die vollständige Menge E von statistischen Einheiten mit gleichen (identisch gleichen) Identifikationsmerkmalen.

Wenn wir im Weiteren einfach von Merkmalen sprechen, dann werden immer die Prädikatsmerkmale der statistischen Einheiten gemeint sein. Je nach den zugrundegelegten Kriterien kann man die Merkmale in verschiedenen Gruppen einteilen. Wir wollen hier eine solche Einteilung behandeln, nämlich die wichtigste und zwar in

1. quantitative Merkmale und
2. qualitative (kategoriale) Merkmale (auch als Attribute bezeichnet).

ad 1. Sie werden gemessen oder gezählt. Bei der Messung wird die Größe des jeweiligen Untersuchungsgegenstandes mit der Größe eines normierten Gegenstandes, der als Einheit dient, verglichen. Quantitative Merkmale sind z.B. Stromverbrauch (gemessen in Kilowattstunden),

Alter (gemessen in Jahren) aber auch wissenschaftliche Mitarbeiter eines Instituts (Anzahl), Wortlänge (Anzahl der Buchstaben). Eine wichtige Unterteilung der quantitativen Merkmale hat Pfanzagl getroffen: die "echten" quantitativen Merkmale zeichnen sich dadurch aus, daß man bei ihnen kardinale Vergleiche anstellen kann, z.B. in der Physik: die Länge, das Gewicht, der elektrische Widerstand usw. Ein Gewicht von 200 kg ist doppelt so groß wie ein Gewicht von 100 kg. Oder in der Linguistik: Das Wort "Linguistik" (10 Buchstaben) ist doppelt so lange wie das Wort "Roman" (5 Buchstaben). Hingegen wäre es sinnlos, einen Studenten mit einer 1 (sehr gut) im Diplom für doppelt so erfolgreich (oder klug oder fleißig) zu halten, wie einen Studenten mit einer 2 (gut) im Diplom. Nach Pfanzagl: "Wir können Skalen, die nur eine Ordnung zum Ausdruck bringen, als *topologische* Skalen bezeichnen; Skalen, deren Informationsgehalt darüber hinausgeht, bei denen es sinnvoll ist, Differenzen zu vergleichen usw., bezeichnen wir als *metrische* Skalen". ⁶⁾

Daraus entsteht die Unterteilung in "topologisch skalierte" und "metrisch skalierte" Merkmale. Bei den ersten sind nur ordinale, bei den zweiten aber kardinale Vergleiche möglich. Die topologische Skalierung sagt nur etwas über die Richtung von Differenzen (von Unterschieden) aus, z.B. "mehr als", "weniger als", "größer als", "kleiner als". Aussagen über die genaue Differenz (über die Größe der Distanz) sind hier nicht möglich. Das ist in einer metrischen Skalierung möglich, die Aussagen der Form z.B. "um soviel mehr", "um soviel weniger", "um soviel größer", "um soviel kleiner" erlaubt.

Die quantitativen Merkmale lassen sich in stetige (kontinuierliche) und diskrete (diskontinuierliche) Merkmale unterteilen.

Von stetigen Merkmalen sprechen wir dann, wenn die Modalitäten jeden Wert innerhalb eines bestimmten Bereichs annehmen können, d.h. wenn sie innerhalb eines bestimmten Intervalls an jeder Stelle der Zahlengerade auftreten oder auftreten können. Diskrete Merkmale zeichnen sich dadurch aus, daß die Modalitäten vereinzelte Werte annehmen (oder auch ganzzahlig auftreten); stetige sind alle metrisch skalierten Merkmale. Zu der ersten Gruppe gehören Merkmale wie Größe, Gewicht, Wachstumsraten usw. Diskret sind Merk-

male wie Zahl der Personen eines Haushalts, Zahl der Bücher einer Bibliothek, Wortlänge (in Buchstaben), Textlänge (in Wörtern) usw. ad 2. Qualitative Merkmale werden durch eine nicht numerische Eigenschaft charakterisiert und sind immer vom diskreten Typ. Für verschiedene Forschungszwecke mag die Zuordnung von Zahlenwerten zu den verschiedenen Modalitäten von qualitativen Merkmalen zweckmäßig sein. Der weitgehend arbiträre Charakter dieser "Quantifizierung" macht deswegen die Mitwirkung des Fachmanns des jeweiligen Wissenschaftszweiges hier besonders erforderlich. Die Menge M der Elementarmodalitäten besitzt hier endlich viele Elemente im Gegensatz zu den quantitativen Merkmalen, deren M auch abzählbar und überabzählbar viele Elemente besitzen kann. Gefragt wird nach der Häufigkeit des Auftretens (oder Nichtauftretens) einer bestimmten Qualität, z.B. Haarfarbe, politische Überzeugung, Wortschatz (nach der Häufigkeit der Verwendung der Wörter Mutter, Vater, haben usw.). Qualitative Merkmale sind per definitionem vom diskreten Typ. Die statistische Theorie wird manchmal nach der eben besprochenen Einteilung der Merkmale differenziert. Statistische Massen von Einheiten mit qualitativen Merkmalen werden in der statistischen Theorie des *homograden* Falles und solche mit quantitativen Merkmalen in der Theorie des *heterograden* Falles behandelt.

2. Empirische Verteilungen

2.1. Organisation statistischer Daten. Häufigkeitsverteilungen.

Bei einer Klausur beteiligten sich 50 Studenten. Die Höchstzahl der zu erreichenden Punktzahl bei richtiger Beantwortung aller gestellten Fragen beträgt 100. Nach der Korrektur der Klausurarbeiten ergaben sich folgende 50 Gesamtpunktezahlen.

Eine andere Fragestellung zu dem gleichen Datenmaterial könnte auch lauten:

Für eine bestimmte linguistische Untersuchung ist es erforderlich,

die empirische Verteilung der in einem fünfzigseitigen Buch vorkommenden bestimmten Artikeln zu untersuchen. Man zählt seitenweise die vorkommenden bestimmten Artikeln und kommt zu folgenden zahlenmäßigen Ergebnissen (oder anders ausgedrückt: zu folgender Anzahl der bestimmten Artikeln, die in den 50 Seiten des untersuchten Buches vorkommen).

Tab. 1

Klausurergebnisse von 50 Studenten
Gesamtpunktezahl der Klausurarbeiten

33	60	74	57	85	52	65	77	84	65
71	81	68	54	35	35	50	74	64	53
41	80	77	45	61	91	55	73	59	47
41	78	76	60	55	69	48	67	85	94
66	94	88	89	98	66	73	42	65	39

Aus dem Urmaterial, wie in Tabelle 1 aufgeführt, lassen sich kaum irgendwelche Aussagen machen. Wir können keine näheren Angaben über die Verteilung der Studenten auf die verschiedenen Klausurleistungen machen. Wir müssen deswegen das Urmaterial so ordnen, daß die Übersichtlichkeit steigt. Zuerst ordnen wir das Urmaterial nach steigenden Zahlen und fassen gleiche Zahlen zusammen, indem wir in einer Spalte aufführen, wie oft diese Zahlen im Urmaterial auftreten (ihre Häufigkeit).

Tab. 2

Klausurergebnisse

Punkte	Anzahl der Studenten f_i	Punkte	f_i	Punkte	f_i	Punkte	f_i
33	1	53	1	67	1	80	1
35	2	54	1	68	1	81	1
39	1	55	2	69	1	84	1
41	2	57	1	71	1	85	2
42	1	59	1	73	2	88	1
45	1	60	2	74	2	89	1
47	1	61	1	76	1	91	1
48	1	64	1	77	2	94	2
50	1	65	3	78	1	98	1
52	1	66	2				
Σ							50

Diese erste Straffung des Urmaterials zur *ungruppierten Häufigkeitsverteilung* ist aber noch nicht befriedigend, denn sie gibt noch keine Auskunft z.B. über die Frage nach der durchschnittlichen Leistung, die charakteristisch für mehr als 3 Studenten sein soll. Bei dieser primitiven Organisation des statistischen Urmaterials gewinnen wir aber schon eine erste charakteristische Kennzahl, die der *Spannweite*, das ist der Bereich, in dem sich die Studentenleistungen bei dieser Klausur bewegen. Sie ist $98-33=65$, d.h. die Differenz zwischen der höchsten und der niedrigsten von einem Studenten erreichte Gesamtpunktzahl.

Der nächste Schritt besteht nunmehr darin, den ganzen Bereich der Spannweite in Unterbereiche zu unterteilen. Wir fassen die geordneten

Beobachtungsdaten in Klassen oder Gruppen zusammen. Die Unterbereiche, die die Klassen bilden, heißen Klassenintervalle. Dabei taucht nun die Frage auf, welche Größe das Klassenintervall haben soll, bzw. wie viele Klassen gebildet werden sollen. Bei gegebenen Beobachtungsdaten bestimmt die Weite des Intervalls die Anzahl der Klassen und umgekehrt. Was wir bei einer Gruppierung des Beobachtungsmaterials erhalten, sind Klassenhäufigkeiten.

Klassenintervalle müssen nicht gleich sein. Es empfiehlt sich aber, ungleiche Intervalle nur dann zu bilden, wenn die Beschaffenheit des Beobachtungsmaterials es erfordert, d.h. wenn dadurch an Übersichtlichkeit gewonnen wird ohne Informationsverlust oder Möglichkeit einer Irreführung.

Welche Weite das Klassenintervall haben soll, bzw. wieviele Klassen gebildet werden, liegt weitgehend im Ermessen der Forschung. Der Forscher muß andererseits bei der Klassenbildung die Struktur der Untersuchungsphänomene und den Typus und die Genauigkeit der vorgenommenen Messungen vor Augen haben.

Wenn wir nun als Klassenintervall einmal 5 und einmal 10 nehmen, erhalten wir folgende Häufigkeitsverteilungen

Klausurergebnisse

Tab. 3 a		Tab. 3 b	
Klasse	f_i	Klasse	f_i
30-34	1	30-39	4
35-39	3	40-49	6
40-44	3	50-59	8
45-49	3	60-69	12
50-54	4	70-79	9
55-59	4	80-89	7
60-64	4	90-99	4
65-69	8		
70-74	5		
75-79	4		
80-84	3		
85-89	4		
90-94	3		
95-99	1		
Σ	50		50

Die Schreibweise 30-34 usw. ist so zu verstehen, daß alle Fälle zu dieser Klasse gehören, die 30 bis *einschließlich* 34 Punkten haben. Eine ähnliche Schreibweise wäre 30 bis *unter* 35. (Diese letzte Schreibweise empfiehlt sich für stetige Merkmale). Wie wir aus diesen beiden Tabellen ersehen, ist die Aussagekraft des Beobachtungsmaterials gestiegen, wenn auch verbunden mit einem gewissen Informationsverlust. Der Statistiker hat also unter Berücksichtigung der Beschaffenheit des Urmaterials die Vorteile einer kleineren Anzahl von Klassen, mit den damit verbundenen Nachteilen des größeren Informations-

verlustes abzuwägen und eine optimale Lösung zu finden. Wenn die Klassenintervalle zu klein für das Beobachtungsmaterial sind, dann werden die Häufigkeiten keinerlei Regularität zeigen; wenn sie zu groß sind, dann wird die jeweilige Klassenhäufigkeit, die dem Mittelwert der Klasse zugerechnet wird, erheblich von der Häufigkeit abweichen, die diesem Wert im Beobachtungsmaterial tatsächlich zukommt (oder wenn dieser Mittelwert bei den Beobachtungsdaten nicht auftritt, von den benachbarten Werten erheblich abweichen) und damit einen bedeutenden Informationsverlust und Verzerrung der Struktur des Beobachtungsmaterials zur Folge haben.

Für längere Beobachtungsreihen hat H. A. Sturges eine aus der Erfahrung abgeleitete Regel, aus der die gesuchte Größe des Intervalls errechnet wird, formuliert:

$$i = \frac{\text{Spannweite}}{1 + 3,322 \cdot \log n}$$

wobei i = das gesuchte Intervall

n = Summe der Häufigkeiten (Anzahl der beobachteten Daten)

Dieser theoretische Wert kann eine Dezimalzahl sein. Dann wird eine runde Zahl in der Nähe des theoretischen Wertes gewählt.

Strauch hat für die maximale Anzahl der Klassen die Bedingung aufgestellt

$$\alpha_{\max} = \frac{\log n}{\log 2} + 1$$

wobei α_{\max} = die maximale Klassenanzahl. Außerdem sollte man darauf achten, daß mindestens im wichtigsten zentralen Bereich der Verteilung keine Leerklassen auftreten, d.h. alle Klassen mit beobachteten Werten besetzt sind. Am besten versucht man, Leerklassen überhaupt zu vermeiden.

Auf unser Beobachtungsmaterial angewendet, errechnet man $i = 9,8$ oder aufgerundet eine Intervallsbreite $i = 10$ und $\alpha_{\max} = 6,6 n$ oder aufgerundet eine maximale Klassenanzahl von $\alpha_{\max} = 7$. Wie man sieht, erfüllt Tab. 3 b beide numerischen Bedingungen: sie weist außerdem keine Leerklassen auf.

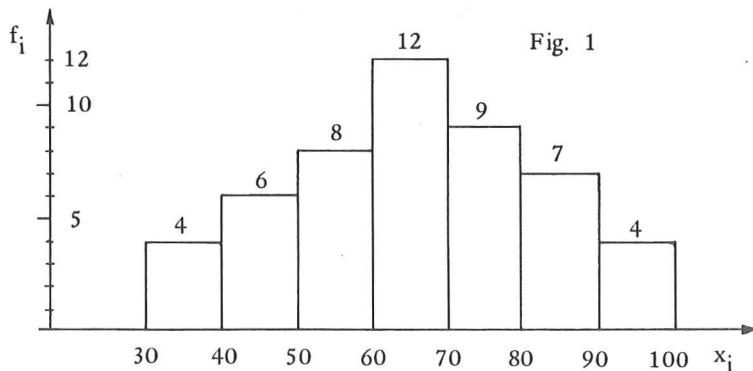
Wenn wir nun die Tabellen 3 a und 3 b vergleichen, sehen wir, daß Tab. 3 b zur Beurteilung der Klausurergebnisse der 50 Studenten größere Aussagekraft besitzt als Tab. 3 a. Sie zeigt größere Regelmäßigkeit, und der Informationsverlust ist verhältnismäßig gering.

2.2. Graphische Darstellung.

Die graphische Darstellung einer Häufigkeitsverteilung informiert sehr oft klarer (weil überschaubarer) und einprägsamer als Häufigkeitstabellen.

(I) Das Histogramm.

In einem rechtwinkligen Koordinatensystem werden die Häufigkeiten auf der Ordinate und die Klassenintervalle auf der Abszisse abgetragen. Jede Klasse bildet eine gleichbreite Säule (bei gleichgroßen Klassenintervallen, sonst unterschiedlich), deren Fläche von der dazugehörigen Häufigkeit der in dieser Klasse zusammengefaßten Merkmalswerte abhängt.



Da wir es hier mit diskreten Merkmalen zu tun haben, müssen wir erst zu dieser kontinuierlichen Darstellung gelangen. Zwischen dem Klassenintervall 20-29 und dem darauffolgenden 30-39 gibt es eine Leerstelle, weil die Bewertung der Klausurergebnisse in ganzen Punktzahlen erfolgt. In solchen Fällen gehen wir so vor, daß wir Kontinuität in der Leerstelle annehmen und den Trennpunkt zwischen beiden Intervallen feststellen, nämlich $(29+30) : 2 = 29,5$. Dadurch vermeiden wir die Leerstellen und setzen die Intervallsgrenzen entsprechend auf diese Trennpunkte.

Bei der Interpretation des Histogramms muß man auf die drei verschiedenen Komponenten der Säulen achten und diese klar erkennen, damit keine Fehler auftreten. Die Höhe der Säule zeigt die Dichte der Häufigkeit pro Klassenintervall. Dies erscheint zwar zuerst wenig verständlich, es wird aber klar, wenn man analoge Begriffsbildungen betrachtet, die geläufig sind. Man sagt z.B., die Bevölkerungsdichte der Bundesrepublik (242 Einwohner pro Quadratkilometer) sei hoch, während die von Frankreich (91 Einwohner pro Quadratkilometer) niedrig (Bezugsjahr 1968) sei.

Die Konzeption der Dichte basiert also in der Bezugnahme auf eine Einheit (pro Einheit). Die Bezugseinheit hier ist das Klassenintervall. Wenn wir in Fig. 1 die vierte Säule betrachten, dann sehen wir, daß die Säule die Höhe 12 hat, was wiederum besagt, daß in dem Intervall 60-69 die Häufigkeitsdichte 12 beträgt, oder anders ausgedrückt: in dem Intervall 60-69 wurden 12 Werte der Variablen x tatsächlich beobachtet. Die Bezugseinheit ist infolgedessen

$$\frac{f}{\text{int}} \quad (\text{Häufigkeit pro Intervall})$$

In dem hier betrachteten Fall muß daher die Höhe der Säule als

$$12 \, f / \text{int}$$

interpretiert werden.

Durch Multiplizieren der Häufigkeitsdichte mit der Länge des Klassenintervalls (Höhe mal Breite) erhalten wir die Fläche der Säule:

$$\left(\frac{12 f}{\text{int}} \right) \times (\text{int}) = 12 f$$

Die Fläche einer Säule zeigt infolgedessen die Häufigkeit in diesem Intervall. In unserer Fig. 1 haben wir gleichgroße Intervalle, sodaß die Häufigkeitsdichte, $12 f/\text{int}$, und die Häufigkeit in diesem Intervall, $12 f$, zahlenmäßig gleich sind. Anders ist es, wenn wir ungleiche Klassenintervalle haben, wie z.B. in der folgenden Tab. 4.

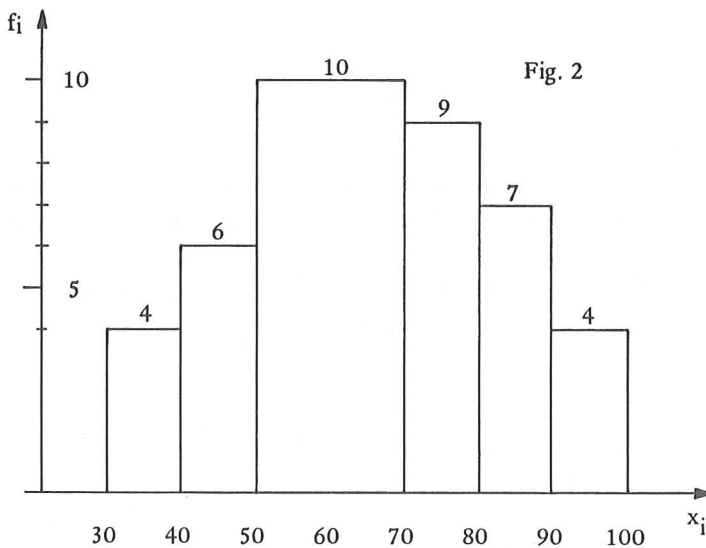
Tab. 4

Klausurergebnisse - ungleiche Klassenintervalle

Klasse	f_i
30-39	4
40-49	6
50-69	20
70-79	9
80-89	7
90-95	4
Σ	50

Das Klassenintervall 50-69 ist doppelt so groß wie die übrigen Intervalle (deren Größe wir als Standardeinheit zugrundelegen), d.h. gleich 2 Standardeinheiten. Die Häufigkeit in diesem Intervall ist 20, d.h. die Fläche der entsprechenden Säule, während die Häufigkeitsdichte, die Höhe der Säule, d.h. die Häufigkeit pro Standardeinheit (pro Standard-

intervall) $20 f / 2 \text{ int} = 10 f$ ist.



In Fig. 2 hat die dritte Säule von links die Häufigkeitsdichte 10, und die Fläche der Säule, die die Häufigkeit in diesem Intervall repräsentiert, ist

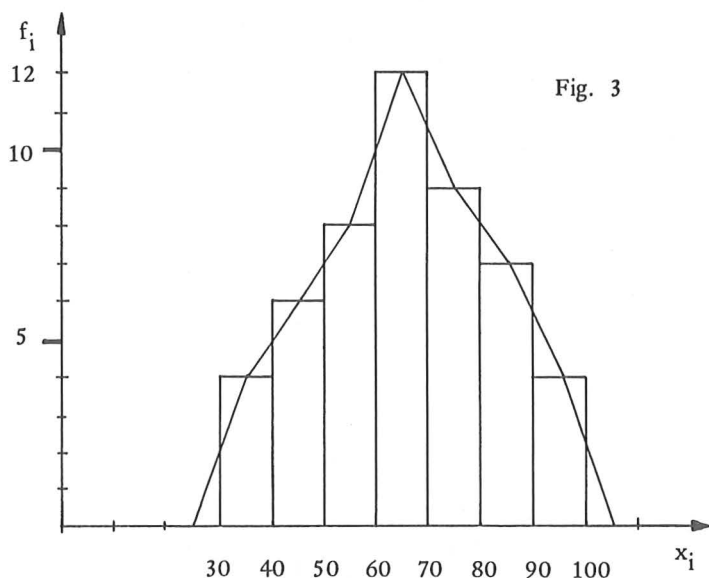
$$\frac{10 f}{\text{int}} \times 2 \text{ int} = 20 f.$$

Die Flächen der Säulen über den verschiedenen Intervallen sind den Häufigkeiten in diesen Intervallen proportional. Die Gesamtfläche aller Säulen repräsentiert die Gesamtzahl aller beobachteten Fälle.

(II) Das Häufigkeitspolygon.

Wenn man in einem Histogramm jeden Klassenmittelpunkt

eines Klassenintervalls mit dem nächsten durch eine Gerade verbindet, erhält man ein Häufigkeitspolygon.



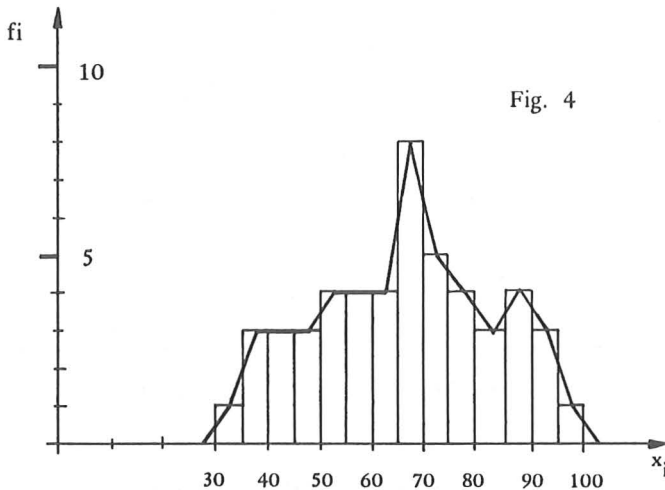
Die zu verbindenden Punkte haben als Abszisse den jeweiligen Klassenmittelpunkt und als Ordinate die Klassenhäufigkeit. Bei der Konstruktion des Häufigkeitspolygons werden außer den aus dem Beobachtungsmaterial gewonnenen Klassen zwei weitere dazu genommen, nämlich eine vor der untersten und eine nach der obersten beobachteten Klasse. In unserem Beispiel sind das die Klassen 20-30 und 100-110. Sie enthalten den Anfang und das Ende der Polygons mit dem jeweiligen Mittelpunkt als Abszisse und mit der Häufigkeit Null. Dadurch wird das Polygon bis zur Abszisse geführt, und die eingeschlossene Fläche repräsentiert die Gesamtzahl der Beobachtungsdaten.

(III) Die Häufigkeitskurve.

Wenn man bei einem Häufigkeitspolygon die Größe der Klasseninter-

valle bis zu einem gewissen Punkt verkleinert, werden die Übergänge von einer Klassensäule zu der nächsten immer kleiner, glatter, gleichförmiger. Von einem gewissen Punkt ab zeigen sich aber Unterbrechungen in den Daten und entsprechend im Diagramm. Die Regularität des Verlaufs der Klassenhäufigkeit, die bei größerem Klassenintervall beobachtet war, wird durch das Auftreten von Klassen, die dieser

5 Punkte Klassenbreite



Regularität zu widersprechen scheinen, unterbrochen. Solche Irregularitäten sind, wie es scheint, Abweichungen von der allgemeinen Regularität, die durch ein Ansteigen der Punktezah von der unteren Grenze bis zu einem Maximalwert und dann durch ein Abnehmen bis zur höchsten Punktzahl charakterisiert ist. Wenn unser Beobachtungsmaterial umfangreicher wäre, z.B. nicht nur eine Klausur, sondern die Klausurergebnisse von einem ganzen Jahr, und/oder nicht nur von 50 Studenten, sondern von allen Studenten einer Universität, dann wären diese Irregularitäten mit zunehmendem Umfang des Beobachtungsmaterials immer geringer. Es ist also wichtig, zu bemerken, daß Be-

wahrung der Regularität und Glättung des Polygons zu einer Kurve nicht nur Verkleinerung des Klassenintervalls, sondern auch Vergrößerung des Beobachtungsmaterials zwecks Ausschaltung akzidenteller Irregularitäten erforderlich macht. Im allgemeinen ist aber eine solche Vergrößerung des Umfangs des Beobachtungsmaterials praktisch unmöglich. Die Häufigkeitskurve wird daher als die graphische Darstellung einer theoretischen Häufigkeitsverteilung angesehen. Darüber wird an anderer Stelle zu sprechen sein.

2.3. Verteilung Relativer Häufigkeiten.

Wenn die Häufigkeiten nicht in absoluten Zahlen, sondern in Anteilswerten an dem Gesamtaufkommen (Prozente oder Bruchteile von 1) angegeben werden, dann spricht man von relativen Häufigkeiten. Wenn wir Tab. 3 b zur Errechnung der relativen Häufigkeiten heranziehen, dann erhalten wir Tab. 5 (relat. Häufigkeit = f').

Tab. 5

Klausurergebnisse

Klasse	f_i	f'_i	
30-39	4	0,08	8 v.H.
40-49	6	0,12	12 v.H.
50-59	8	0,16	16 v.H.
60-69	12	0,24	24 v.H.
70-79	9	0,18	18 v.H.
80-89	7	0,14	14 v.H.
90-99	4	0,08	8 v.H.
Σ	50	1,00	100 v.H.

Wenn man zwei Häufigkeitsverteilungen vergleichen will, deren zugrundegelegtes Beobachtungsmaterial größenmäßig sehr verschieden ist, dann ist es angebracht, die relativen Häufigkeiten zu verwenden.

Bei der Aufstellung einer Verteilung der relativen Häufigkeiten ist es zweckmäßig, die Gesamtzahl der beobachteten Fälle (Summe der absoluten Häufigkeiten) anzugeben, denn dadurch erhält man die Möglichkeit, durch Multiplikation dieser Summe mit den einzelnen relativen Häufigkeiten auf die jeweiligen absoluten zu gelangen.

Hinsichtlich der graphischen Darstellung gilt auch hier, *mutatis mutandis*, das bei der Verteilung der absoluten Häufigkeiten gesagte.

2.4. Kumulative Häufigkeitsverteilung.

Wenn wir bei unserem Beispiel der Klausurergebnisse die Frage stellen, wieviele Studenten eine Gesamtpunktzahl von über 50 (mehr als 50) erreicht haben, dann ist es notwendig, das Beobachtungsmaterial anders aufzubereiten und graphisch darzustellen. Das gleiche gilt, wenn die Frage umgekehrt formuliert ist, nämlich wieviele Studenten weniger als 70 Punkte erreicht haben.

Bei der Beantwortung dieser Fragen wird es notwendig sein, eine kumulative Häufigkeitsverteilung aufzustellen. Sie gibt an, wie viele Beobachtungen auf einem variablen Wert und unterhalb bzw. oberhalb von ihm liegen. Die erste Kumulierungsart nennt man Aufwärts- die zweite Abwärtskumulation; diese Bezeichnungen werden weiter unten durch die Tabulierung und die graphische Darstellung einsichtig werden.

Tab. 6

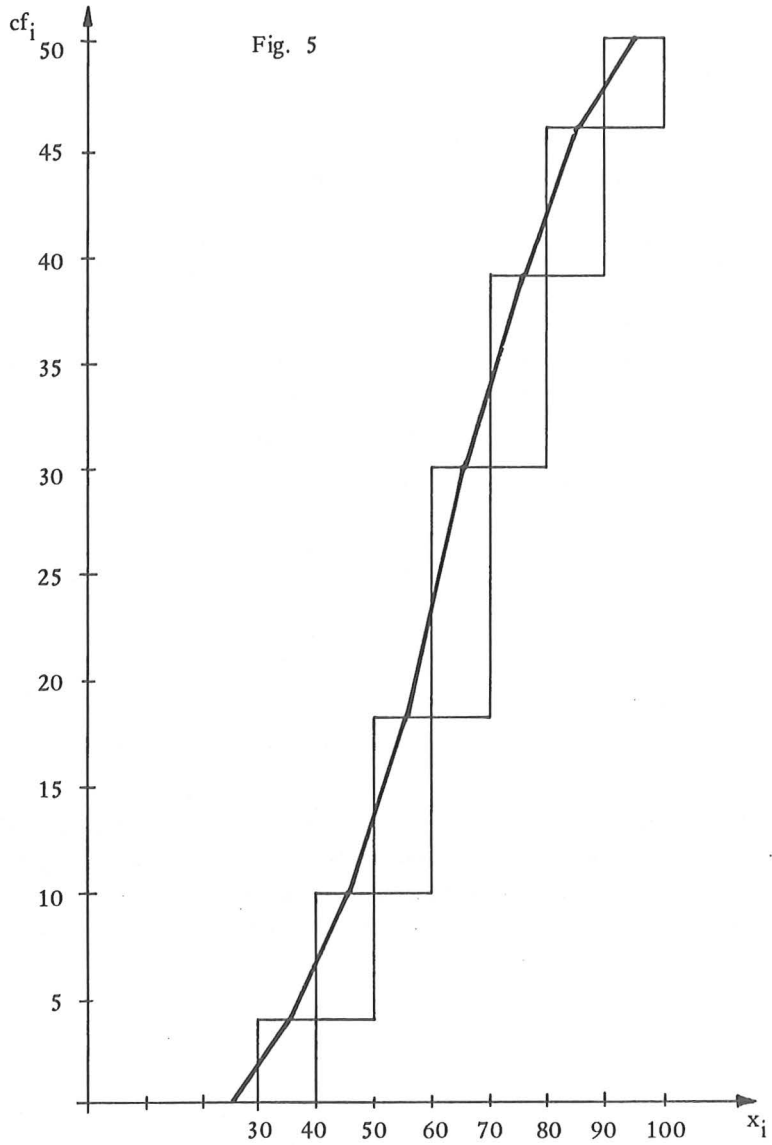
Kumulative Häufigkeitsverteilung der
Klausurergebnisse von 50 Studenten

- Aufwärtskumulation -			- Abwärtskumulation -		
x_i (1)	cf_i (2)	cf'_i (3)	x_i (4)	cf_i (5)	cf'_i (6)
unter 29	0	0,00	über 30	50	1,00
" 39	4	0,08	" 40	46	0,92
" 49	10	0,20	" 50	40	0,80
" 59	18	0,36	" 60	32	0,64
" 69	30	0,60	" 70	20	0,40
" 79	39	0,78	" 80	11	0,22
" 89	46	0,92	" 90	4	0,08
" 99	50	1,00	" 100	0	0,00

Die kumulative Häufigkeitsverteilung kann durch ein Summenpolygon (Ogive) graphisch veranschaulicht werden. (Aus dem Summenpolygon entsteht durch Glättung die Summenkurve oder Kumulativkurve). Die Ogive ist ein sehr leistungsfähiges Instrument zur Repräsentation von Häufigkeitsreihen.

Ogive Aufwärtskumulation

Fig. 5



Sie hat den großen Vorteil, daß sie prinzipiell unabhängig von den zugrundegelegten Klassenintervallen und der Anzahl der Klassen ist, d.h. die Summenkurve bleibt grundsätzlich gleich, auch bei Variation der Klassenintervalle. Unterschiedliche Klassenintervalle beeinträchtigen nicht die Gestalt der Ogive, so daß auch in einem solchen Fall Ogiven vergleichbar bleiben, Histogramme dagegen nicht.

Darüberhinaus ist die Ogive für Interpolationen geeignet. (Unter Interpolation versteht man die Ermittlung einer Zahl, die zwar im Beobachtungsmaterial oder in der aufbereiteten Tabelle nicht auftritt, innerhalb der gegebenen Spannweite aber vorhanden ist). Die technische Möglichkeit der Interpolation entbindet natürlich, wie so oft in der Statistik, den Wissenschaftler nicht von der Aufgabe, über Sinn und Aussagekraft einer solchen Interpolation nachzudenken und wenn notwendig, diesen Schritt vom Untersuchungsgegenstand her zu begründen und zu interpretieren.

Figur 5 ist aus Figur 1 bzw. Tab. 3b, d.h. aus dem Histogramm, entstanden, insofern gilt auch hier, was wir dort über die Häufigkeiten gesagt haben.

Aus den oben erwähnten Eigenschaften der Ogive ergibt sich aber weiterhin, daß sie auch unmittelbar aus dem geordneten Urmaterial gewonnen werden kann.

In Spalte (3) und (6) der Tab. 6 sind auch die kumulativen relativen Häufigkeiten errechnet; sie zeigen keine besonderen Schwierigkeiten und werden deswegen nicht weiter diskutiert.

Wenn man beide Kumulationsarten auf das gleiche Koordinatensystem einträgt, dann werden sich die zwei Ogiven an einem Punkt schneiden, der genau die Hälfte der beobachteten Fälle (50%) repräsentiert.

2.5. Qualitative Merkmale.

Bei der bisherigen Erörterung der Häufigkeitsverteilungen haben wir uns mit quantitativen Merkmalen beschäftigt. Bei den qualitativen Merkmalen wird nach dem Auftreten oder Nicht-Auftreten einer bestimmten Qualität gefragt und die entsprechende Häufigkeit gezählt.

Die graphische Darstellung kann in diesem Fall nicht mehr in der Form eines Histogramms, einer Häufigkeitskurve oder einer Ogive erfolgen, was aus sachlogischen Gründen, d.h. aus dem Charakter der qualitativen Merkmale unmittelbar einsichtig ist.

Die adäquate graphische Darstellung für qualitative Merkmale ist das Stabdiagramm.

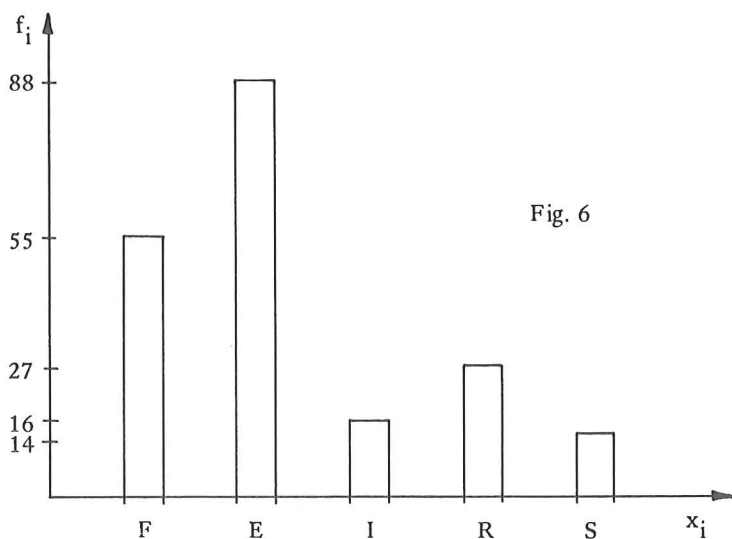
Die beobachtete Häufigkeit (absolut oder relativ) des jeweiligen qualitativen Merkmals wird durch die Höhe des Stabes repräsentiert. Breite und Fläche des Stabes haben, da wir es hier nicht mit Häufigkeitsdichte zu tun haben, keine Bedeutung, ebenso wenig die Lage der Stäbe entlang der Koordinatenachse.

In einer Universität wurden die 200 Studenten, die Fremdsprachen studieren, nach ihrem Hauptfach klassifiziert. Daraus ergab sich folgende Tabelle:

Tab. 7

Sprache	Anzahl der Studenten
Französisch	55
Englisch	88
Italienisch	16
Russisch	27
Spanisch	14

Die graphische Darstellung in einem Stabdiagramm ist in Fig. 6 ausgeführt.



2.6. Die Lorenz-Kurve.

Wenn wir unsere statistische Reihe Tab. 3 b, bzw. 5 in Anteilswerten an dem jeweiligen Gesamtaufkommen ausdrücken, dann erhalten wir folgende Tabelle

Tab. 8 a				Tab. 8 b		
Klasse	f_i	x'_i	f'_i		$c.x'_i$	$c.f'_i$
				unter 29	0,00	0,00
30-39	4	0,04	0,08	" 39	0,04	0,08
40-49	6	0,08	0,12	" 49	0,12	0,20
50-59	8	0,13	0,16	" 59	0,25	0,36
60-69	12	0,24	0,24	" 69	0,49	0,60
70-79	9	0,21	0,18	" 79	0,70	0,78
80-89	7	0,18	0,14	" 89	0,88	0,92
90-99	4	0,12	0,08	" 98	1,00	1,00
Σ	50	1,00	1,00			

In Tab. 8 a sind sowohl die Häufigkeiten als auch die Klassifizierungsbasis in relativen Werten ausgedrückt worden. In Tab. 8 b sind diese relativen Werte aufwärts kumuliert worden. Wenn wir diese Werte in ein Koordinatensystem eintragen, ergibt sich die in Fig. 7 dargestellte Graphik: Kurve OBA.

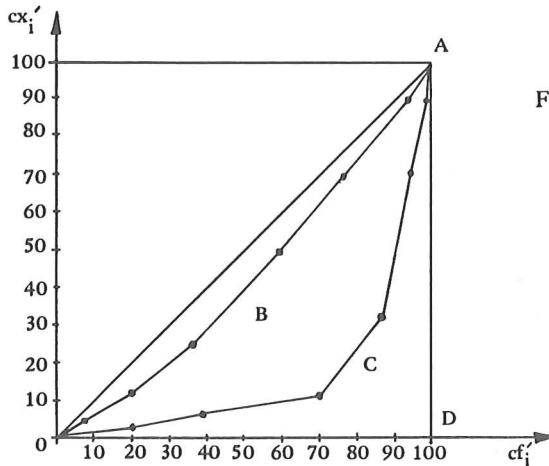


Fig. 7

Wie man aus der Graphik entnehmen kann, ist eine absolut gleiche Verteilung der Werte der Ordinate, sagen wir der bestimmten Artikel, durch die Linie OA repräsentiert. Sie ist eine Gerade und bildet einen Winkel von 45 Grad. (Der Maßstab sollte natürlich für beide Achsen gleich sein, was aus dem Charakter beider Reihen, in relativen kumulativen Werten ausgedrückt - unmittelbar nicht nur möglich ist, sondern auch für angebracht erscheint).

Bei einer solchen Verteilung würden beispielsweise 20 % der untersuchten Textseiten 20 % der vorkommenden bestimmten Artikeln, 30 % der Seiten 30 % der bestimmten Artikeln usw. enthalten.

Je größer die Abweichung von der Gleichverteilung ist (d.h. je größer die Konzentration der Häufigkeitsverteilung), desto größer wird die Abweichung der Kurve der kumulativen, relativen Häufigkeitsverteilung von der Linie der Gleichverteilung OA sein.

Die Linie ODA zeigt eine absolut ungleiche Verteilung.

Wenn wir statt der in Tab. 8 a bzw. 8 b angegebenen Verteilung die in Tab. 9 hätten, würden wir die Kurve

Tab. 9

x_i'	f_i'	cx_i'	cf_i'
		0,00	
0,03	0,20	0,03	0,20
0,02	0,18	0,05	0,38
0,05	0,32	0,10	0,70
0,20	0,20	0,30	0,90
0,40	0,06	0,70	0,96
0,18	0,02	0,88	0,98
0,12	0,02	1,00	1,00

OCA erhalten, die weiter von der Kurve der Gleichverteilung OA entfernt ist als OBA.

Die Lorenz-Kurve bildet ein geeignetes Instrument zum anschaulichen Vergleich zweier oder mehrerer statistischer Reihen und zwar hinsichtlich ihrer Konzentration. Dabei ist auch ein effektiver Vergleich der Konzentrationsgrade zu unterschiedlichen Zeitpunkten (soweit natürlich der Zeitfaktor eine Rolle spielt) oder unter verschiedenen Bedingungen möglich.

3. Maßzahlen der Häufigkeitsverteilung

Die im vorigen Kapitel behandelten Methoden der Organisierung und Darstellung von statistischen Daten bilden eine erste Stufe der Beschreibung, eine Vorbereitung zur eigentlichen Beschreibung des Beobachtungsmaterials. Diese erste Stufe erschöpft sich aber in einer tabellarischen oder graphischen Organisierung des statistischen Materials. Die Beschreibung der Häufigkeitsverteilung ist unzureichend und macht exakte quantitative Vergleiche größerer Aussagekraft kaum möglich. Eine Häufigkeitsverteilung läßt sich durch verschiedene Zahlen charakterisieren, die spezifische Besonderheiten eines statistischen Beobachtungsmaterials beschreiben. Diese statistische *Maßzahlen* oder *Parameter* haben den Vorteil, daß sie die ganze Vielfalt der Details des statistischen Materials auf eine einzelne Zahl reduzieren und darüber hinaus rechnerische Operationen und quantitative Vergleiche erlauben.

Wir unterscheiden 4 solche charakteristische Maßzahlen der Häufigkeitsverteilung; sie beschreiben und messen verschiedene signifikante Eigenschaften einer statistischen Masse: (1) Die zentrale Tendenz, (2) die Streuung, (3) die Schiefe und (4) die Wölbung einer Häufigkeitsverteilung.

3.1. Maßzahlen der zentralen Tendenz.

Diese Maßzahlen werden Mittelwerte (oft auch wenig präzise Durchschnitts) genannt. Die Aussagekraft, d.h. der repräsentative Charakter der Mittelwerte, hängt von einem Mindestmaß an Konzentration und Variabilität (des zu repräsentierenden Beobachtungsmaterials) ab. Wenn wir beispielsweise die Beobachtungsdaten 8, 220, 1500, haben, dann ist das arithmetische Mittel von 576 von sehr beschränkter Aussagekraft. Ähnliches gilt vom Mittelwert 6 gleichgroßer Zahlen.

3.1.1. Das arithmetische Mittel.

Dieser wichtigste und gebräuchlichste aller Mittelwerte entsteht durch

Summierung aller Beobachtungswerte, dividiert durch die Anzahl der in ihr enthaltenen Beobachtungen. In mathematischer Schreibweise drückt sich das so aus:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

wobei n die Anzahl der Beobachtungen und x_i der Wert der einzelnen Beobachtungen sind.

Bei gruppierten Beobachtungen, d.h. Häufigkeitsverteilungen, wird das arithmetische Mittel durch die Formel

$$\bar{x} = \frac{\sum_{i=1}^k x_i \cdot f_i}{\sum_{i=1}^k f_i}$$

errechnet, wobei f_i die Häufigkeit der einzelnen Klassen und die Summe aller Häufigkeiten gleich der Anzahl der Beobachtungen n sind. Bei den gruppierten Häufigkeitsverteilungen entsteht allerdings die Frage, welcher Wert für x_i eingesetzt werden soll. Bei dem von uns verwendeten Beispiel würde also die Frage zu beantworten sein, welchem Wert z.B. die Häufigkeit 8 der Klasse 50-59 zuzurechnen wäre. Man geht hier so vor, daß man den Mittelpunkt des Klassenintervalls m_i ermittelt und diesen Wert als adäquate Repräsentation des mittleren Wertes dieser Klasse annimmt. Diese Annahme setzt wiederum die Annahme

voraus, daß sich die in jeder Klasse beobachteten Häufigkeiten gleichmäßig über die jeweilige Klasse verteilen. Bei großen beobachteten Häufigkeitszahlen ist diese Annahme gewöhnlich eine gute Approximation.

Bei ungruppierten Häufigkeitsverteilungen sind diese m_i gleich den Merkmalswerten x_i .

In der Klasse 50-59 haben wir aus Tab. 2 die tatsächlichen Werte $50 + 52 + 53 + 54 + 55 + 55 + 57 + 59 = 435$. Wenn wir diese Summe durch die Anzahl der Beobachtungen, 8, dividieren,

Tab. 10

Klausurergebnisse von 50 Studenten

Klasse	f_i	m_i	$f_i m_i$
30-39	4	34,5	138,0
40-49	6	44,5	267,0
50-59	8	54,5	436,0
60-69	12	64,5	774,0
70-79	9	74,5	670,5
80-89	7	84,5	591,5
90-99	4	94,5	378,0
Σ	50		3255,0

erhalten wir : $435 : 8 = 54,375$. Wie man sieht, ist die Abweichung des aktuellen Mittelwertes vom Klassenmittelpunkt so gering, nämlich 0,125 , daß man von einer sehr guten Approximation sprechen kann.

Das arithmetische Mittel wird nunmehr nach der modifizierten Formel

$$\bar{x} = \frac{\sum_{i=1}^n m_i \cdot f_i}{\sum_{i=1}^n f_i}$$

errechnet.

Da die Häufigkeit die Wichtigkeit, das Gewicht der einzelnen Merkmalswerte in der Häufigkeitsverteilung, anzeigt, spricht man in diesem Fall von einem gewogenen arithmetischen Mittel: die Merkmalswerte werden mit der Häufigkeit ihres Auftretens "gewogen" oder "gewichtet".

Das arithmetische Mittel kann auf einem Punkt liegen, an dem relativ oder absolut wenige oder sogar keine Beobachtungen auftreten (oder für den letzten Fall: keine Beobachtungen auftreten können). In unserem Beispiel ist $\bar{x} = 65,1$ und das ist eine Merkmalsausprägung, die wegen der Struktur des Beobachtungsmaterials nicht auftreten kann.

3.1.2. Der Median, (Zentralwert).

Der Median teilt eine Häufigkeitsverteilung oder eine ungruppierte Datenreihe in zwei gleiche Teile, so daß die Hälfte der beobachteten Merkmalsausprägungen oberhalb und die andere Hälfte unterhalb von ihm liegen. Das ist recht einfach bei ungerader Zahl der Beobachtungen. Wenn wir beispielsweise 51 geordnete Beobachtungen haben, dann ist der 26. Beobachtungswert der Median mit 25 Beobachtungswerten über und 25 unter ihm. Daraus ersehen wir, daß der Median für ungruppierte, geordnete Beobachtungsreihen durch die Formel

$$\bar{x}_z = \frac{n + 1}{2}$$

errechnet werden kann.

Etwas komplizierter wird es bei gruppierten Häufigkeitsverteilungen. Die Berechnung des genauen Wertes des Medians kann dann nach der Formel erfolgen:

$$\bar{x}_z = G + i \cdot \frac{\frac{n}{2} - f_v}{f_e}$$

wobei die Symbole folgendermaßen zu interpretieren sind:

G = unterer Grenzwert der Klasse, in der der Median liegt (Einfallsklasse)

n = Gesamtzahl der Beobachtungen

f_e = Besetzungszahl (Häufigkeit) der Einfallsklasse

f_v = Besetzungszahl aller der Einfallsklasse vorausgehender Klassen.

i = Klassenintervall.

Nach dieser Formel errechnen wir den Median unserer Häufigkeitsverteilung der Tab. 7 und erhalten:

$$\bar{x}_z = 60 + 10 \cdot \frac{25 - 18}{12} = 65,83 \approx 66$$

Er beträgt also 66 Punkte (oder genau 65,83).

Dabei ist die Annahme zugrundegelegt worden, daß die einzelnen Elemente einer Klasse innerhalb dieser Klasse gleichmäßig verteilt sind.

Der Median reagiert nicht wie das arithmetische Mittel auf alle Änderungen der einzelnen Merkmalswerte (wobei das arithmetische Mittel nur auf eine einzige Änderung nicht reagiert, nämlich auf eine, die zwar einzelne Merkmalswerte ändert, diese Änderungen aber sich in soweit gegenseitig aufheben, daß das Gewicht der beiden Abschnitte der Verteilung im physikalisch-metaphorischen Sinne unverändert bleibt und infolgedessen der Gleichgewichtspunkt auch).

Zur Veranschaulichung dieser einfachen Eigenschaft nehmen wir ein einfaches Beispiel: wir haben die Merkmalswerte 5, 7, 12, 21, 26; dann ist 12 der Median und 14,2 das arithmetische Mittel.

Wenn wir jetzt die Merkmalswerte 1, 4, 12, 15, 19 nehmen, dann ist 12 wieder der Median, während das arithmetische Mittel 10,2 beträgt.

3.1.3. Der Modus (Häufigster Wert).

Der Modus ist der Merkmalswert mit der größten Häufigkeitskonzentration und hat infolgedessen nur bei einer Häufigkeitsverteilung eine Bedeutung. Er ist (nach einem Ausdruck von Fechner) der dichteste Wert einer Verteilung.

Die Berechnung des Modus einer Häufigkeitsverteilung ist nicht immer leicht und eindeutig durchzuführen. Unterschiedliche Klassenein-

teilungen, d.h. Klassenintervalle und Klassengrenzwerte, können zu verschiedenen Werten des Modus führen. Bei theoretischen Häufigkeitsverteilungen ist eine exakte Bestimmung des modalen Wertes allerdings möglich.

Für praktische Zwecke genügt allerdings eine approximative Bestimmung des Modus. Wenn die allgemeinen Regeln der Gruppierung beachtet werden, dann wird auch der Fehler des errechneten Modalwertes klein bleiben.

Die Ermittlung des Modus kann dann durch verschiedene Annäherungsformel erfolgen:

$$\bar{x}_m = G + i \cdot \frac{f_a}{f_a + f_b}$$

wobei

f_a = Besetzungszahl (Häufigkeit) der der Modalklasse vorhergehenden Klasse

f_b = Besetzungszahl der der Modalklasse nachfolgenden Klasse.

Daraus können wir den Modus für unsere Verteilung ermitteln, und wir erhalten als Wert $\bar{x}_m = 64,70$.

Eine andere Annäherungsformel sieht folgendermaßen aus:

$$\bar{x}_m = G + i \cdot \frac{\Delta a}{\Delta a + \Delta b}$$

wobei Δa = Differenz zwischen der Besetzungszahl der Modalklasse und der der Modalklasse direkt vorhergehenden Klasse = $f_c - f_a$

Δb = Differenz zwischen der Besetzungszahl der Modal-
 klasse und der der Modalklasse direkt nachfolgenden
 Klasse = $f_e - f_b$

Diese Formel kann man infolgedessen auch in folgender Gestalt
 schreiben:

$$\bar{x}_m = G + i \cdot \frac{f_e - f_a}{2f_e - f_a - f_b}$$

Der daraus ermittelte Modus beträgt $x_m = 65,71$.

Eine genauere Berechnung des Modus kann durch folgenden Rechen-
 gang erreicht werden:

Wir gehen von der Überlegung aus, daß die einzelnen Beobachtungs-
 werte innerhalb der Einfallsklasse des Modus eine ähnliche Vertei-
 lung aufweisen wie diejenigen, außerhalb dieser Klasse. In unserem
 Fall liegen 18 Beobachtungswerte unterhalb und 20 oberhalb der Ein-
 fallsklasse 60 - 69. Die proportionale Aufteilung der Modalklasse
 hat also nach den Verhältnissen $\frac{18}{38} : \frac{20}{38} = 0,48 : 0,52$

zu erfolgen. Der Modus wird näher an der oberen Grenze liegen als
 an der unteren. Wir teilen das Klassenintervall nach diesem Verhält-
 nis $0,48 \times 10 = 4,80$ und $0,52 \times 10 = 5,20$, und erhalten einen
 Modus von $60 + 5,20 = 65,20$. Wenn wir das in Form einer Formel
 bringen wollen, erhalten wir zwei Fälle:

$$(1) \quad \text{Wenn } \frac{f_v}{n - f_e} > \frac{f_n}{n - f_e} \quad \text{dann}$$

$$\bar{x}_m = G + i \cdot \frac{f_v}{n - f_e}$$

$$(2) \text{ wenn } \frac{f_n}{n - f_e} > \frac{f_v}{n - f_e} \quad \text{dann}$$

$$\bar{x}_m = G + i \frac{f_n}{n - f_e} \quad \bullet$$

3.1.4 Charakterische Eigenschaften der wichtigsten Mittelwerte.

Arithmetisches Mittel

Vorteile	Nachteile
<ol style="list-style-type: none">1. Das arithmetische Mittel ist leicht zu berechnen.2. Es ist ein rechnerisches Mittel und daher für mathematische Manipulationen geeignet.3. Es ist in jedem Fall eindeutig bestimmt.4. Es schließt alle Beobachtungswerte einer Reihe ein und wird von jedem einzelnen Wert beeinflusst.5. Es weist eine größere Stichprobenstabilität aus als die anderen Mittelwerte.	<ol style="list-style-type: none">1. Extreme und oft atypische Werte erhalten bei der Berechnung des arithmetrischen Mittels zu großes Gewicht.2. Infolge seiner Beeinflußbarkeit von Extremwerten kann das arithmetische Mittel auf einem Punkt liegen, in dem wenige oder gar keine Beobachtungswerte vorliegen; es ist bei unsymmetrischen Verteilungen am weitesten vom evidenten Konzentrationspunkt der Verteilung entfernt.

Median

<ol style="list-style-type: none">1. Er ist leicht verständlich.2. Er wird von Extremwerten nicht tangiert.3. Er kann auch bei unvollständigen Datenreihen ermittelt werden, wenn die	<ol style="list-style-type: none">1. Sehr geringe mathematische Manipulierbarkeit.2. Wenn Extremwerte einer Datenreihe als typische charakteristische Bestandteile einer Datenreihe betrachtet
---	---

<p>Anzahl der Beobachtungen und genaue Informationen zum mittleren Bereich der Reihe bekannt sind.</p> <p>4. Bei topologisch skalierten Merkmalen ist der Median den anderen Mittelwerten überlegen wegen seiner Invarianz gegenüber monotonen Transformationen. "Der Median der transformierten Werte ist gleich dem transformierten Median der ursprünglichen Werte" ⁷⁾.</p>	<p>werden, dann ist der Median dem arithmetischen Mittel unterlegen, weil diese Werte bei der Ermittlung des Medians kein Gewicht haben.</p>
---	--

Modus

<p>1. Er wird von extremen Werten nicht beeinflusst.</p> <p>2. Der approximative häufigste Wert ist leicht zu ermitteln.</p> <p>3. Der Modus ist der typischste Mittelwert einer Häufigkeitsverteilung, da er an dem Punkt der größten Konzentration liegt.</p>	<p>1. Die Ermittlung des genauen Modus erfordert bei Häufigkeitsverteilungen umfangreichere Rechnungen.</p> <p>2. Der Modus hat keine Bedeutung, wenn die Häufigkeitsverteilung eine kleinere Anzahl von Beobachtungen enthält und keine ausgesprochene zentrale Tendenz aufweist.</p> <p>3. Er ist mathematisch in sehr geringem Maße manipulierbar.</p>
---	---

3.2. Maßzahlen der Streuung

Die im letzten Abschnitt besprochenen Mittelwerte beschreiben eine Häufigkeitsverteilung nur unvollkommen. Denn ein Mittelwert allein besitzt eine kleine Aussagekraft, solange man über die Abweichungen der einzelnen Beobachtungswerte einer statistischen Reihe von einem gewonnenen Mittelwert keine Information besitzt. Wenn wir nicht wissen, wie die einzelnen Beobachtungswerte gestreut sind, d.h. welchen Variabilitätsgrad die Häufigkeitsverteilung der Beobachtungswerte aufweist, können wir keine hinreichende Beschreibung einer solchen Verteilung geben. Bei unserem Klausurbeispiel würde z.B. der Vergleich unserer o.a. Ergebnisse mit den Ergebnissen einer anderen ungefähr gleich großen Klasse zu keiner wichtigen Aussage führen können, wenn die beiden Datenreihen einen ungefähr gleichen Mittelwert hätten. Die eine Reihe könnte dann z.B. dadurch charakterisiert sein, daß die Mehrzahl der Studenten dicht beieinander liegende Gesamtpunktezahlen erzielt haben, während die andere Reihe viele unterschiedliche Gesamtpunktezahlen der Studenten aufweist. Der ungefähr gleiche Mittelwert würde in diesem Fall ein falsches, weil unvollständiges Bild von der Struktur der Häufigkeitsverteilung der Klausurergebnisse liefern.

Wenn wir als zweites Beispiel die Häufigkeitsverteilungen zweier Klausurergebnisse nehmen, und ihre arithmetischen Mittel errechnen, sagen wir 65, dann würde die Schlußfolgerung, daß beide Klassen ungefähr gleich gut abgeschnitten haben, ein Schuß ins Blaue sein: denn die eine Reihe könnte Gesamtpunktezahlen umfassen, die sich von 15 bis 95 bewegen, während die Ergebnisse der anderen Reihe in der Spannweite von 40 bis 97 liegen könnten.

Im folgenden werden vier Maßzahlen der Streuung behandelt: die Spannweite, die mittlere lineare Abweichung, die mittlere quadratische Abweichung und der Variationskoeffizient.

3.2.1. Spannweite.

Wir haben diese Maßzahl schon kennengelernt.

3.2.2. Die mittlere lineare Abweichung.

Die mittlere Abweichung einer statistischen Reihe ist das arithmetische Mittel der absoluten Abweichungen der einzelnen Beobachtungswerte von einem Mittelwert, gewöhnlich vom Median oder vom arithmetischen Mittel. Daraus ergibt sich die entsprechende Formel:

$$MA = \frac{\sum_{i=1}^k f_i |x_i - \bar{x}|}{n}$$

Bei Häufigkeitsverteilungen verwenden wir als x_i den jeweiligen Klassenmittelpunkt. Für unsere Verteilung der Klausurergebnisse berechnen wir die mittlere Abweichung nach folgender Tabelle

Tab. 11

Arbeitstabelle - mittlere lineare Abweichung

Klassen	x_i	f_i	$ x_i - \bar{x} $	$f_i \cdot x_i - \bar{x} $
30-39	34,5	4	30,6	122,4
40-49	44,5	6	20,6	123,6
50-59	54,5	8	10,6	84,8
60-69	64,5	12	0,6	7,2
70-79	74,5	9	9,4	84,6
80-85	84,5	7	19,4	135,8
90-99	94,5	4	29,4	117,6
Σ		50		676,0

Wenn wir in die Formel einsetzen, erhalten wir

$$MA = \frac{676}{50} = 13,52$$

Die mittlere Abweichung wird selten benutzt, vor allem, wenn das statistische Material in Form einer Häufigkeitsverteilung organisiert ist. Sie wird nur bei statistischen Reihen kleineren Umfangs benutzt, wenn keine exakte und detailliert ausgearbeitete Analyse des statistischen Materials angestrebt wird. Das ist hauptsächlich auf die logischen und mathematischen Begrenzungen dieser Maßzahl zurückzuführen.

3.2.3. Die mittlere quadratische Abweichung (Standardabweichung).

Die wichtigste und am häufigsten verwendete Maßzahl der Streuung ist die Standardabweichung. Mathematisch gesehen ist die Standardabweichung definiert als die Quadratwurzel aus dem arithmetischen Mittel der quadrierten Abweichungen (Differenzen) der Einzelwerte einer statistischen Reihe von ihrem arithmetischen Mittel. Als Mittelwert, auf den die Abweichungen der Einzelwerte bezogen werden, nimmt man also das arithmetische Mittel, und zwar deswegen, weil die Summe der quadrierten Abweichungen vom Mittel kleinsten sind, kleiner als die Abweichungen von irgendeinem anderen Wert. (vergl. Exkurs 1).

Die Standardabweichung wird mit dem griechischen Buchstaben sigma dargestellt.

$$\sigma = \frac{\sum_{i=1}^k f_i (x_i - \bar{x})^2}{n}$$

Da die rechnerische Arbeit zur Ermittlung der Standardabweichung ziemlich umfangreich ist, werden sehr oft vereinfachte Rechenverfahren verwendet. Es gibt verschiedene solche Verfahren; alle stützen sich aber auf zwei wichtige Eigenschaften der Standardabweichung.

Erste Eigenschaft. Wenn jedem Beobachtungswert einer statistischen Reihe eine konstante Zahl addiert wird, dann bleibt σ unverändert.

Betrachten wir den Ausdruck $(x_i - \bar{x})^2$ in der obigen Formel. Addieren wir zu jedem Beobachtungswert x_i die Konstante c , dann ändert sich auch das arithmetische Mittel um diese Konstante. Der Ausdruck erhält dann die Form $[(x_i + c) - (\bar{x} + c)]^2$
 $= (x_i + c - \bar{x} - c)^2 = (x_i - \bar{x})^2$.

Daraus folgt, daß die Standardabweichung unabhängig von der numerischen Größe der Beobachtungswerte der statistischen Reihe ist. Der numerische Wert der Standardabweichung wird durch die relativen Distanzen der Beobachtungswerte untereinander bestimmt.

Zweite Eigenschaft. Wenn jeder Beobachtungswert einer statistischen Reihe mit einer Konstanten multipliziert wird, dann ändert sich σ um den gleichen multiplikativen Faktor.

Multiplizieren wir die Formel für σ mit dem konstanten Faktor c , dann ergibt sich

$$\sqrt{\frac{\sum_{i=1}^n (cx_i - c\bar{x})^2}{n}} = \sqrt{\frac{c^2 \sum_{i=1}^n (x_i - \bar{x})^2}{n}} = c \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = c\sigma$$

Die Standardabweichung ändert sich nicht aufgrund der Änderung in der Größe der statistischen Daten, sondern infolge der Veränderungen der Distanzen der Daten untereinander. Die Streuung wird größer, wenn die multiplikative Konstante c größer als 1 ist, und sie wird kleiner, wenn der multiplikative Faktor eine Dezimalzahl unter 1 ist.

Wenn wir jetzt die Standardabweichung unserer Häufigkeitsverteilung errechnen wollen, erhalten wir folgende Arbeitstabelle

Tab. 12

Arbeitstabelle - Standardabweichung

Klassen	x_i	f_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
30-39	34,5	4	-30,6	936,36	3745,44
40-49	44,5	6	-20,6	424,36	2546,16
50-59	54,5	8	-10,6	112,36	898,88
60-69	64,5	12	- 0,6	0,36	4,32
70-79	74,5	9	9,4	88,36	795,24
80-89	84,5	7	19,4	376,36	2634,52
90-99	94,5	4	29,4	864,36	3457,44
Σ		50			14082,00

Wir setzen jetzt in die Formel ein, und erhalten

$$\sigma = \sqrt{\frac{14082}{50}} = \sqrt{281,64} = 16,782$$

Unter Bezugnahme auf die zwei Eigenschaften der Standardabweichung können wir ein vereinfachtes Rechenverfahren zur Ermittlung von σ anführen. Dieses Rechenverfahren findet bei gruppierten statistischen Reihen, d.h. bei gruppierten Häufigkeitsverteilungen, Anwendung. In einem ersten Schritt addiert man irgendeine konstante Zahl c_1 (gewöhnlich mit negativem Vorzeichen) zu den einzelnen Daten, so daß man kleinere Zahlen erhält, dann multipliziert man die nunmehr modifizierten Einzeldaten mit einer anderen Kon-

stanten c_2 zwecks weiterer Vereinfachung der Datenreihe. Die aus dieser modifizierten Zahlenreihe erhaltene Standardabweichung wird dann mit $1/c_2$ multipliziert und man erhält dadurch die Standard-

abweichung der ursprünglichen Reihe. Wir wenden dieses Verfahren auf unsere Häufigkeitsverteilung an und erhalten folgende Arbeitstabelle.

Als c_1 wird meistens ein zentraler Wert genommen, weil dadurch die Differenzen nach beiden Seiten gleichmäßig klein werden. Oft wird der Klassenmittelpunkt der modalen Klasse genommen, weil dadurch die Differenz 0 wird und die Multiplikation mit einer großen Häufigkeitszahl zu keiner großen Zahl führt. Als c_2 wird oft die Größe des Klassenintervalls genommen. Es ist aber bei jeder Häufigkeitsverteilung von neuem zu überlegen, wie ihre Datenzahlen und die durchzuführenden Operationen am besten vereinfacht werden können.

Tab. 13

Arbeitstabelle - Standardabweichung

$c_1 = 64,5$ Punkte

$c_2 = 10$ Punkte

Klassen	x_i	f_i	$(x_i - c_1) = d_i$	$\frac{d_i}{c_2} = y_i$	$f_i y_i$	y_i^2	$f_i y_i^2$
30-39	34,5	4	-30	-3	-12	9	36
40-49	44,5	6	-20	-2	-12	4	24
50-59	54,5	8	10	-1	- 8	1	8
60-69	64,5	12	0	0	0	0	0
70-79	74,5	9	10	1	9	1	9
80-89	84,5	7	20	2	14	4	28
99-99	94,5	4	30	3	12	9	36
Σ		50			3		141

Die Formel für die vereinfachte Berechnung der Standardabweichung lautet:

$$\sigma = c_2 \sqrt{\frac{\sum_{i=1}^k f_i y_i^2 - \frac{\left(\sum_{i=1}^k f_i y_i\right)^2}{n}}{n}}$$

Wir setzen die entsprechenden Zahlen aus der Arbeitstabelle ein und erhalten:

$$\sigma = 10 \sqrt{141 - \frac{(3)^2}{50}} = 10 \sqrt{2,816} = 16,78$$

Zum gleichen Ergebnis wären wir gekommen, wenn wir für c_1 den Wert 74,5 genommen hätten. Die Formel für die vereinfachte Berechnung der Standardabweichung bei ungruppierten Verteilungen lautet:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2}$$

Bei statistischen Untersuchungen wird sehr oft vom Quadrat der Standardabweichung Gebrauch gemacht. Diese Größe σ^2 wird Varianz genannt.

Da die Differenzen $(x_i - \bar{x})$ quadriert werden, ist die Summe der quadrierten Abweichungen der einzelnen Beobachtungswerte vom arithmetischen Mittel eine positive Zahl, ihr arithmetisches Mittel auch. Je größer diese Abweichungen sind, desto größer ist die

Standardabweichung. Wenn alle Beobachtungswerte gleich sind, dann ist die Standardabweichung Null.

3.2.4 Der Variations-Koeffizient.

Diese Maßzahl der Streuung basiert nicht wie die drei bis jetzt besprochenen Maßzahlen auf den absoluten Größen der statistischen Reihe und folglich auch nicht auf den absoluten Distanzen, sondern auf der relativen Streuung der Beobachtungsdaten. Wenn wir neben unseren Klausurergebnissen die Klausurergebnisse der gleichen 50 Studenten haben, allerdings in einem anderen Fach, bei dem eine maximale Gesamtpunktzahl von 500 möglich war, dann haben wir bei der ersten Klausur ein $\bar{x}_1 = 65,1$ und $\sigma_1 = 16,78$ und bei der zweiten ein $\bar{x}_2 = 200$ und $\sigma_2 = 22,5$. Absolut gesehen ist die Streuung, die durch σ_2 repräsentiert wird größer als die Streuung die durch σ_1 repräsentiert wird. Ein Vergleich der Variabilität der zwei Ergebnisse ist, wie man hier sieht, nur sinnvoll unter Bezugnahme auf die arithmetischen Mittel, von welchen die Abweichungen gemessen werden. Eine solche Maßzahl ist der Variationskoeffizient einer Verteilung. Er ist definiert als

$$V = \frac{\sigma}{\bar{x}}$$

Er kann auch als die Standardabweichung in Prozent des arithmetischen Mittels ausgedrückt werden, d.h.

$$V = \frac{\sigma}{\bar{x}} \cdot 100$$

Wenn wir nun unsere zwei Klausurergebnisse vergleichen, erhalten wir

$$V_1 = \frac{16,78}{65,1} = 0,288 \text{ oder } 28,80 \%$$

und

$$V_2 = \frac{22,50}{200} = 0,1125 \text{ oder } 11,25 \%$$

Wir ersehen daraus, daß die relative Streuung der Ergebnisse der 2. Klausur erheblich niedriger ist als die relative Streuung der Ergebnisse der 1. Klausur.

Der Variationskoeffizient wäre nicht interpretierbar, wenn das arithmetische Mittel mit dem Koordinatenursprung zusammenfallen würde, d.h. wenn $\bar{x} = 0$ ist.

V würde in diesem Fall für alle Werte der Standardabweichung (außer $\sigma = 0$) gleich unendlich sein.

3.2.5 Der Yule-Herdan-Index.

Die Stabilität der Häufigkeitsparameter arithmetisches Mittel und Standardabweichung (folglich auch Variationskoeffizient) ist bei manchen linguistischen Fragestellung nicht gewährleistet. Das ist nicht auf die grundsätzliche Unzulänglichkeit dieser Parameter, sondern auf spezifische, dem allgemeinen Sprachsystem immanente Eigenschaften zurückzuführen. Hier zeigt sich wiederum die Notwendigkeit einer ständigen Beschäftigung des Fachwissenschaftlers, hier des Linguisten, mit dem Instrumentarium der Statistik im Lichte der theoretischen Grundlagen und der objektiven Gegebenheiten des Untersuchungsgegenstandes seiner wissenschaftlichen Disziplin. In der linguistisch-statistischen Charakterisierung der Worthäufigkeitsverteilung eines Textes erweisen sich die vorgenannten Parameter als nicht sehr leistungsfähig, denn die Verteilung dieser linguistischen Einheiten unter Zugrundelegung der Häufigkeit ihres Auftretens ist nicht unabhängig von der jeweiligen Textlänge (L). Daher ist auch ein Vergleich verschieden langer Texte (L_1, L_2, \dots ; $L_1 \neq L_2$ usw.) hinsichtlich des Wortgebrauchs unter Verwendung dieser Parameter nicht ohne weiteres möglich.

Da die Anzahl der Wörter, die einem bestimmten Autor zur potentiellen Aktualisierung im Rahmen eines Textes (streng genommen auch

in einem bestimmten Zeithorizont) zur Verfügung steht, sich in engen Grenzen hält, wird sich die Vergrößerung der Textlänge ($L_2 > L_1$, d.h. statistisch die Vergrößerung der Ereigniszahl, wobei

$$L = \sum_{i=1}^k f_i x_i$$

ist) in einer Vergrößerung der Häufigkeit der Verwendung (des Auftretens) einzelner Wörter auswirken.

Aus diesen Überlegungen läßt sich die Versuchsanordnung zur Untersuchung der Worthäufigkeitsverteilung folgendermaßen beschreiben:

Die Merkmalausprägung des zu untersuchenden linguistischen Materials wird durch die jeweilige zahlenmäßige Aktualisierungshöhe eines Wortes in einem Text oder Textausschnitt festgelegt.

Die spezifische Charakterisierung eines Wortes in einem bestimmten Text wird also durch eine Wiederholungsrate x_i gegeben. Die Anzahl der Wörter, die in einem bestimmten Text oder Textausschnitt x_i -mal auftreten, d.h. diese Merkmalausprägung aufweisen, ist die Häufigkeit f_i dieses Ereignisses.

Aus dieser Versuchsanordnung ergeben sich folgende wichtige Größen:

$$\sum_{i=1}^k f_i = N \quad \text{d.h. die Wortliste}$$

$$\sum_{i=1}^k f_i x_i = L \quad \text{d.h. die Gesamtzahl der Ereignisse oder die Textlänge in Wörtern gemessen.}$$

Wenn wir nun einen Text von L Wörtern Länge haben mit der Liste der verschiedenen ⁸⁾ Wörtern N , wobei $L > 1$

und $N > 1$, dann können wir theoretisch zwei Fälle unterscheiden:

(1) $L = N$, d.h. alle Wörter in diesem Text sind verschieden.

(2) $L > N$, d.h. einige Wörter werden wiederholt.

Der Fall (2) charakterisiert hauptsächlich die in der Realität vorkommenden Texte ⁹⁾.

Aus diesen Überlegungen resultiert ein Ansteigen des Wertes sowohl des arithmetischen Mittels als Verhältniszahl zwischen der

$$\text{Gesamtzahl der Ereignisse zu der Wortliste} \quad \frac{\sum_{i=1}^k f_i \cdot x_i}{\sum_{i=1}^k f_i} \quad \text{als auch}$$

des Wertes der Standardabweichung mit der Länge des Textes, da die Häufigkeit der Verwendung der schon eingetretenen Wörter schneller ansteigt als die Zunahme der Anzahl der aktualisierten Vokabulareinheiten, d.h. anders ausgedrückt, die Textlänge wächst schneller als die Wortliste. (In einer logischen a priori Bedingungskonstellation von einer bestimmten "Länge" des "Textes" ab).

Mathematisch ausgedrückt: $\Delta L > \Delta N$ oder $\frac{\Delta L}{\Delta N} > 1$.

Zur Überwindung dieser Instabilität der herkömmlichen Parameter und zur Erreichung einer parametrischen Vergleichbarkeit verschieden langer Texte, gemäß der durch die beschriebene Versuchsanordnung gestellten Fragestellungen, sind verschiedene neue Maßzahlen vorgeschlagen worden ¹⁰⁾.

Als bester Parameter im Sinne der Fragestellung hat sich die "Charakteristik" K von Yule erwiesen. Die Ableitung von K erfolgt aber unter Zugrundelegung einer Poisson-Verteilung der Worthäufigkeiten. Yule betont jedoch: "... the complete accident-distribution could be described by a binomial distribution with negative index. The two parameters of which could be simply in terms of its mean and standard deviation. I made one or two attempts to see whether this expression could be made to describe a word-distribution with any accuracy ... but they were all complete failures",¹¹⁾

Herdan hat diesen Parameter ohne Rückgriff auf die für diese Situation inadäquate Poisson-Verteilung abgeleitet,¹²⁾ in dem er den quadrierten Variationskoeffizienten V durch N dividiert:

$$\begin{aligned} \frac{V^2}{N} &= \frac{\sigma^2/N}{\bar{x}^2} = \frac{\frac{1}{N} \left(\frac{\sum_{i=1}^k f_i x_i^2}{N} - \frac{\left(\sum_{i=1}^k f_i x_i \right)^2}{N^2} \right)}{\frac{\left(\sum_{i=1}^k f_i x_i \right)^2}{N^2}} = \\ &= \frac{\frac{1}{N^2} \left(\sum_{i=1}^k f_i x_i^2 - \frac{\left(\sum_{i=1}^k f_i x_i \right)^2}{N} \right)}{\frac{\left(\sum_{i=1}^k f_i x_i \right)^2}{N^2}} = \frac{\sum_{i=1}^k f_i x_i^2}{\left(\sum_{i=1}^k f_i x_i \right)^2} - \frac{1}{N} \end{aligned}$$

Für großes N kann aber $\frac{1}{N}$ vernachlässigt werden.

Wir erhalten daher

$$V^2_{\bar{x}} = \frac{\sum_{i=1}^k f_i x_i^2}{\left(\sum_{i=1}^k f_i x_i \right)^2}$$

Statistisch stellt dieser Parameter den Variationskoeffizienten des arithmetischen Mittel dar. Wie wir sehen, ist dieser Parameter von der Größe von N, d.h. vom Umfang der Stichprobe ¹³⁾ unabhängig.

Verschiedene statistische Untersuchungen haben gezeigt, daß diese Formel weitgehend stabil bleibt, während die anderen Parameter eine systematische Änderung mit dem Stichprobenumfang erfahren ¹⁴⁾

Diese Stabilität des Yule-Herdan-Parameters

$$V_{\bar{x}} = \frac{\sigma / \sqrt{N}}{\bar{x}}$$

für homogenes Material erlaubt es, ihn zur Charakterisierung der spezifischen quantitativen Eigenschaften eines Textes hinsichtlich der Worthäufigkeitsverteilung zu verwenden. ¹⁵⁾ Nach Herdan "The interpretation of V_m ($V_{\bar{x}}$ in unserer Terminologie) is that it represents a measure of concentration of vocabulary on comparatively few vocabulary items. The greater V_m , the greater the contrast between the number of words used only once and those used a very great number of times; the smaller V_m , the more even will be the distribution of the occurrence frequencies over the different vocabulary". ¹⁶⁾

3.3 Maßzahlen der Schiefe.

Die bisherigen Maßzahlen gaben uns Auskunft über die zentrale Tendenz einer Häufigkeitsverteilung und über den Konzentrationsgrad

(das Ausmaß der Streuung) um diese zentrale Tendenz. Damit ist aber eine statistische Datenreihe noch keineswegs vollständig beschrieben. Eine weitere Beschreibungskomponente bilden die Maßzahlen der Schiefe, die den Grad der Asymmetrie, d.h. die Abweichung einer gegebenen Verteilung von einer symmetrischen Verteilung messen.

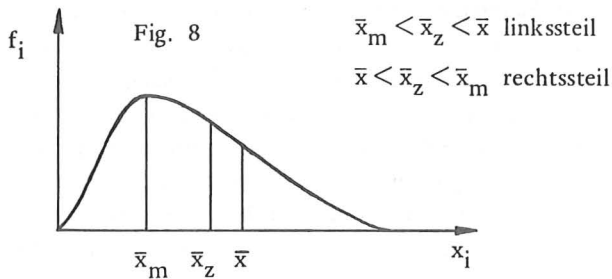
Ein sehr einfaches Maß für die Schiefe ist die Differenz zwischen arithmetischem Mittel und Modus, bezogen auf die Standardabweichung:

$$sk = \frac{\bar{x} - \bar{x}_m}{\sigma}$$

oder die Differenz zwischen arithmetischem Mittel und Median, bezogen auf die Standardabweichung:

$$sk = \frac{3(\bar{x} - \bar{x}_z)}{\sigma} \quad (\text{sk vom englischen Skewness})$$

Diese Maßzahl kann positive und negative Werte annehmen, je nachdem ob das arithmetische Mittel größer als der dichteste Wert, bzw. als der Zentralwert ist, oder kleiner. Dies zeigt wiederum, ob die Verteilung linkssteil oder rechtssteil ist, also die Richtung



und das Ausmaß der Asymmetrie. Die zweite Formel kann nur bei Verteilungen mit ausgeprägter Asymmetrie angewandt werden. Beide Maßzahlen können bei bestimmten Beschaffenheiten des statistischen Materials unbrauchbar werden.

Wir geben hier zwei andere Maßzahlen der Schiefe, die zu einem späteren Zeitpunkt weiter erläutert werden. Der Koeffizient von Pearson:

$$\beta_1 = \frac{\left[\frac{1}{n} \sum_{i=1}^k f_i (x_i - \bar{x})^3 \right]^2}{\sigma^2}$$

bei $\beta_1 \neq 0$ liegt Asymmetrie der Verteilung vor. Die Richtung der Asymmetrie läßt sich aber nicht feststellen, da durch die Potenzierung mit einer geraden Zahl immer positive Ergebnisse erzielt werden.

Der Koeffizient von Fisher

$$\gamma_1 = \frac{\frac{1}{n} \sum_{i=1}^k f_i (x_i - \bar{x})^3}{\sigma^3}$$

Dieser Koeffizient gibt auch die Richtung der Asymmetrie an, ist aber nicht von der gleichen Schärfe wie β_1 .

3.4 Maßzahlen der Wölbung

Sie geben die Abweichungen einer gegebenen Verteilung von der Normalverteilung, die in dem Teil der induktiven Statistik behandelt wird, an, und zwar in der Höhe, in der Wölbung, auch Kyrstosis genannt.

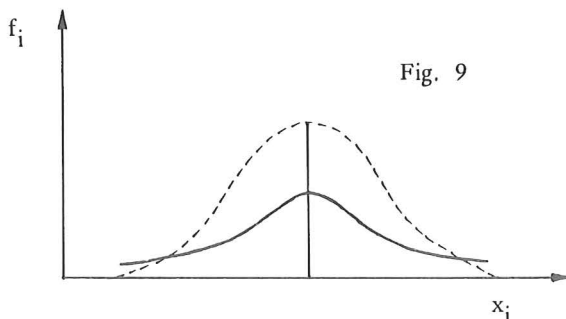


Fig. 9

In Fig. 8 haben die zwei Verteilungen sowohl gleiche Maßzahlen der zentralen Tendenz als auch gleiche Streuung und Schiefe; sie haben aber unterschiedliche Wölbung.

Wir haben auch hier zwei Koeffizienten.

Koeffizient von Pearson:

$$\beta_2 = \frac{\frac{1}{n} \sum_{i=1}^k f_i (x_i - \bar{x})^4}{\sigma^4}$$

Koeffizient von Fisher:

$$\gamma_2 = \beta_2 - 3$$

Je nach dem Wert der Koeffizienten nennt man eine Verteilung

bei $\beta_2 < 3$ oder $\gamma_2 < 0$ platykurtisch

bei $\beta_2 = 3$ oder $\gamma_2 = 0$ mesokurtisch und

bei $\beta_2 > 3$ oder $\gamma_2 > 0$ leptokurtisch

Die Maßzahlen der Schiefe und der Wölbung werden in der Praxis wenig benutzt, sind aber zur vollständigen Beschreibung einer Häufigkeitsverteilung und zum Vergleich zweier oder mehrerer Verteilungen notwendig.

Exkurs 1

Die Summe der Abweichungen der Einzelwerte vom arithmetischen Mittel ist gleich 0.

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x}) &= \sum_{i=1}^n x_i - n\bar{x} = n \frac{\sum_{i=1}^n x_i}{n} - n\bar{x} \\ &= n\bar{x} - n\bar{x} = 0\end{aligned}$$

Wenn wir irgendeine Konstante c nehmen und das arithmetische Mittel \bar{x} einer statistischen Reihe, dann ist $c - \bar{x} = a$ auch konstant.

$$\begin{aligned}\sum_{i=1}^n (x_i - c)^2 &= \sum_{i=1}^n (x_i - \bar{x} - a)^2 = \sum_{i=1}^n [(x_i - \bar{x}) - a]^2 \\ &= \sum_{i=1}^n \left[(x_i - \bar{x})^2 - 2a(x_i - \bar{x}) + a^2 \right] \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 - 2a \sum_{i=1}^n (x_i - \bar{x}) + na^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(c - \bar{x})^2\end{aligned}$$

Die Summe $(x_i - c)^2$ erhält den kleinsten Wert bei $c - \bar{x} = 0$, d.h. wenn $c = \bar{x}$ ist. Die Summe der quadrierten Abweichungen der Einzelwerte einer Reihe vom arithmetischen Mittel ist ein Minimum im Vergleich zu den Abweichungen von allen anderen Werten (auch Mittelwerten).

C. Messung und Skalierung

*“Sie könnens so ausdrücken:
Wo nichts am rechten Platz
liegt, da ist Unordnung. Wo
am rechten Platz nichts liegt,
ist Ordnung.”*

B. Brecht

1. Allgemeine Charakterisierungen

Es ist natürlich, daß in denjenigen Wissenschaftszweigen, in welchen Messungen unsicher, schwierig und sehr aufwendig sind, die Statistik eine große Rolle spielt. Wo keinerlei Messung möglich wäre, würde keine Anwendung der Statistik möglich sein; dort andererseits, wo jede Untersuchung präzise Messungen erreichen könnte, würde auf manchen, und zwar vielen Gebieten, die Notwendigkeit, Statistik anzuwenden, verschwinden. Zwischen diesen zwei Extremen, hier keine Messung, dort perfekte Messung, erreicht “the ratio of statistizing to measuring” sein Maximum.¹⁷⁾ Damit ist aber auch gesagt daß eine wichtige Voraussetzung irgendeiner Relevanz der statistischen Methodik eine noch im einzelnen genauer zu definierenden Meßbarkeit des Objektbereichs ist, der jeweils untersucht wird.

Unter Messung im weitesten Sinn wollen wir jeden Prozess der Zuordnung von *Zeichen* aus einem konventionell geordneten Zeichensystem zu Objekten, Phänomenen oder Ereignissen gemäß bestimmter Regeln verstehen. (Im weiteren werden wir uns auf das geläufige Zahlensystem beschränken.) Die *Zahl*, die jedem einzelnen Objekt des durch diesen Zuordnungsprozess erfaßten Objektbereichs zugeschrieben wird, nennen wir *Maßzahl*. Den Ausgangspunkt dieses Zuordnungsprozesses bildet eine Eigenschaft oder Erscheinungsform der Objekte oder Ereignisse, die durch das Untersuchungsziel festgesetzt wird; wir nennen sie *Größe*. Die Maßzahl gibt dann eine spe-

zielle Ausprägung dieser Größe wieder, z.B. den Betrag oder den Grad dieser Größe. Hier soll auf die Korrespondenz dieser Definitionen zu den an anderer Stelle getroffenen Feststellungen hinsichtlich der Merkmale und Merkmalsmodalitäten einer statistischen Einheit hingewiesen werden.

Diesen Zuordnungsprozess kann man auffassen als eine Abbildung von Objekten oder Ereignissen in einen abstrakten mathematischen Raum bestimmter Struktur.

Wenn wir eine empirische Menge M haben, deren Elemente durch die Relationen B_1, B_2, \dots, B_n geordnet sind, dann erhalten wir ein empirisches Strukturdatum

$$\mathcal{M} = (M; B_1, B_2, \dots, B_n)$$

nach Tarski auch empirisches relationales System genannt. Durch eine Zuordnungsregel ϕ wird diesem System

$$\mathcal{N} = (N; R_1, R_2, \dots, R_n)$$

zugeordnet, wobei N eine Teilmenge der reellen Zahlen ist. Durch das geordnete Tripel

$$\langle \mathcal{M}, \mathcal{N}, \phi \rangle$$

wird dann eine Skala definiert. "Immer wenn diese Reihenfolge gegeben ist . . . haben wir eine Repräsentation von empirischen Eigenschaften, Strukturdaten als numerische Eigenschaften, Strukturdaten vorliegen. Damit wird sichergestellt, daß die Zuordnung von Zahlen zu Phänomenen keine einmalige ist".¹⁸⁾

Die Messung einer Menge linguistischer (soziologischer, psychologischer, usw.) Erscheinungen besteht also in der Widerspiegelung empirischer Relationen in numerischen im Rahmen einer linguistischen Strukturbeschreibung. Unter welcher Regel diese Widerspiegelung erfolgt, hängt natürlich von der jeweiligen Struktur des empirischen Objektbereichs ab, d.h. aber *auch* von dem jeweils erreichten wissenschaftlichen Stand der Strukturierung dieses Bereichs.

Aus der Tatsache, daß die Zuordnung von Zahlen zu Objekten oder Ereignissen unter verschiedenen Regeln stattfinden kann, folgt, daß wir verschiedene Skalen und verschiedene Messungsverfahren haben.

Eine Skala bringt das Niveau, die Stufe der Meßbarkeit (oder der Quantifizierung, um einen in der Literatur gängigen Ausdruck zu verwenden) zum Ausdruck. Sie legt auch fest, nach welchen Prinzipien ein Maß für eine in einer Untersuchung zugrundegelegte Größe bestimmt werden kann.¹⁹⁾

Der Komplex der Messung und Skalierung läßt sich somit im Rahmen der statistischen Argumentation unter drei Aspekten untersuchen, mit dem Ziel, sie explizit anzugeben. Folgende Aspekte müssen expliziert werden:

- a) die verschiedenen Zuordnungsregeln,
- b) die mathematischen Eigenschaften der sich ergebenden Skalen und
- c) die statistischen Operationen, die bei Messungen mit den einzelnen Skalenarten anwendbar sind.

Wir wollen die ersten zwei Aspekte gemeinsam behandeln und vier verschiedene Skalen beschreiben, die am häufigsten auftreten.

Bei der Entwicklung der verschiedenen Skalenarten muß man darauf achten, daß sie aufeinander aufbauen, d.h. jede weitere Skala "versucht dem Merkmal, um das es geht, auch jene Eigenschaften der Zahlen zuzuschreiben, welche die in der Reihenfolge vorangehende Skala zuordnet. Die Skalen bilden auf Grund der Eigenschaften, die sie zuordnen, eine kumulative Hierarchie".²⁰⁾

2. Skalenarten

2.1. Die Nominalskala

Die Nominalskala repräsentiert die unterste Meßbarkeitsstufe und es wird oft angezweifelt, ob man in diesem Fall überhaupt von Meßbarkeit sprechen kann. Die dem Objektbereich zugeordneten Zahlen fungieren als Namen, wobei die einzigen Aussagen, die man durch Widerspiegelung des empirischen Strukturdatums in das numerische

relationale System machen kann, die der Identität oder Verschiedenheit sind. Wir können zwei verschiedene Typen der Zuordnung unterscheiden, (a) Zuordnung von Zahlen zu einzelnen Objekten oder Ereignissen zwecks Identifikation, (b) Zuordnung von Zahlen zu Klassen von Objekten.

Typ (a) ist natürlich ein Spezialfall von Typ (b), denn wir können die einzelnen Objekte als Klassen mit einem Element ansehen.

Die Nominalskala stellt die einfachste Form einer Strukturierung des Objektbereichs dar. Durch die Klassifikation und auf der Grundlage einer gemeinsamen Eigenschaft oder Erscheinungsform wird ein Objektbereich in Klassen aufgeteilt, d.h. die Objekte eines Bereichs werden auf Grund eines Merkmals, das sie von anderen Objekten unterscheidet (auch distinctive feature bezeichnet), verschiedenen Klassen zugeordnet. Eine solche Klassifikation hat folgende formallogische Bedingungen zu erfüllen:

- (1) Die durch die Aufteilung des Objektbereichs gewonnenen Klassen müssen sich wechselseitig ausschließen, d.h. die Klassenextensionen dürfen sich nicht überschneiden.
- (2) Die Aufteilung des Objektbereichs muß erschöpfend sein, d.h. jedes Objekt muß einer Klasse zugeordnet sein.

Mengentheoretisch lassen sich diese Bedingungen folgendermaßen schreiben: wenn M der Objektbereich und

$$M = \{K_1, K_2, \dots, K_n\}$$

die Klasseneinteilung ist, dann muß

- (i) $K_1 \cup K_2 \cup \dots \cup K_n = M$
- (ii) $K_i \cap K_j = \phi \quad (1 \leq i, j, \leq n) \text{ oder}$
 $K_1 \cap K_2 \cap \dots \cap K_n = \phi \text{ sein.}$

“Despite all the differences that exist between the separate Schools in modern linguistics, there is one thing that unites all their tendencies, namely, the method of distributive analysis, which consists in exhibiting classes of elements which are interchangeable in some sense or other. The procedure for verifying the interchangeability, known in the Copenhagen School as ‘test of commutation’, and in Descriptive Linguistics as ‘substitution’, is, it seems, one of the basic instruments of investigation in modern linguistics”. ²¹⁾

Aus den bisherigen Ausführungen geht hervor, daß man durch Äquivalenzrelationen zwischen Elementen des Objektbereichs (aufgedeckt und aufgestellt unter Zugrundelegung einer bestimmten interessierenden Eigenschaft) die Einteilung in Klassen vornehmen kann.

Äquivalenzrelationen müssen aber drei Bedingungen erfüllen:

- E_1 : Für alle $a \in k$ gilt : $a \sim a$ (Reflexivität)
 E_2 : Für alle $a, b \in k$ gilt : $a \sim b \rightarrow b \sim a$ (Symmetrie)
 E_3 : Für alle $a, b, c \in k$ gilt : $a \sim b \wedge b \sim c \rightarrow a \sim c$
 (Transitivität)

Dabei bedeutet $a \sim b$: “a äquivalent b” oder “a ranggleich b” oder “a gleich b” u.ä.m. ²²⁾

Ein bekanntes, anschauliches Beispiel aus der Linguistik, das die Stellung der Transitivitätsbedingung bei der Klassenbildung anzeigt geben G.E. Peterson und Frank Harary, ²³⁾ und zwar im Hinblick auf “semantically equivalent utterances”: “It is possible, of course, that a single utterance may belong to more than one semantic equivalence class, as in the case of homonyms. For example, consider the three utterances x, y and z:

- x. The rays of the sun meet.
- y. The sun’s rays meet.
The sons raise meat.
- z. Meat is raised by the sons.

If we denote the relation of semantic equivalence by E, it can be seen that E is reflexive and symmetric but not transitive. For xEx , yEy etc.;

and if xEy , than yEx , but the above well known example shows that if xEy and yEz , it does not follow that xEz ".

2.2. Die Ordinalskala

Eine höhere Stufe der Meßbarkeit wird durch die Ordinalskala erreicht. Sie entsteht aus der Operation des Vergleichs und der Rangordnung (Reihenfolge) der Elemente einer Menge. Wenn wir eine Menge M mit den Elementen a, b, c, \dots , haben, dann müssen die einzelnen Elemente paarweise in einer Relation zueinander stehen, d.h. die Menge muß geordnet sein. Damit ist natürlich für die einzelwissenschaftliche Praxis die Aufgabe verbunden, in den jeweils untersuchten Objektbereichen, d.h. in empirisch gegebenen Mengen Ordnungen durch paarweisen Vergleich der Elemente aufzudecken - soweit sie vorhanden sind. ²⁴⁾

Die der Axiomatik der Ordinalskala zugrundeliegende Relation ist die der Verschiedenheit, wobei auch die Richtung dieser Relation und der relative Platz in der Rangfolge angegeben werden. Wir wollen diese Relationen mit $a \succ b$ oder $b \prec a$ darstellen, wobei dann der Ausdruck interpretiert werden kann als "a wird b vorgezogen" oder "a ist ranghöher als b" oder "a dominiert b" oder "a ist größer als b" u.ä.m. Die arithmetischen $>, <$ Beziehungen werden dann bei der Widerspiegelung des empirischen in das numerische relationale System die Repräsentation des primär empirischen Sachverhalts übernehmen.

Folgende Eigenschaften dieser Relation müssen erfüllt sein:

- O_1 : Für alle $a, b \in M$ gilt : $a \succ b \rightarrow \neg (b \succ a)$ (Asymmetrie)
- O_2 : Für alle $a, b, c \in M$ gilt : $a \succ b \wedge b \succ c \rightarrow a \succ c$ (Transitivität)
- O_3 : Für alle $a, b \in M$ gilt : eine und nur eine der drei Relationen $a \succ b, b \prec a, a \sim b$ (Trichotomie)

Eine durch R geordnete Menge M (wobei R für \succ, \prec, \sim steht) ist auf einer ordinalen Skala meßbar, wenn eine numerische Funktion $f(x), x \in M$, existiert, so daß für $a, b \in M$ gilt:

$$f(a) \geq f(b) \iff aRb$$

Jede Funktion $f(x)$, die diese Bedingung erfüllt, ist eine ordinale Messung von M , wenn $f(x)$ eine ordinale Messung von M ist und wenn $\hat{\varphi}(t)$ eine wachsende Funktion von t ist, dann ist auch $\hat{\varphi}(f)$ eine ordinale Messung von M ; d.h. ordinale Messungen sind indifferent gegenüber monotonen Skalentransformationen (m.a.w. $x' = f(x)$, wobei $f(x)$ eine monoton wachsende Funktion ist). Eine asymmetrische Relation ist die empirische Beziehung "a ist Ehemann von b"; denn wenn a der Ehemann von b ist, kann b nicht der Ehemann von a sein. Dagegen ist die Beziehung "a ist verheiratet mit b" symmetrisch. Transitiv ist die Relation "a ist Vorfahre von b". Denn aus den Relationen "a ist Vorfahre von b" und "b ist Vorfahre von c" folgt, daß auch "a Vorfahre von c" ist.

Die ordinale Messung erlaubt also eine relative Rangordnung der Elemente des Objektbereichs hinsichtlich einer bestimmten, interessierenden Eigenschaft. Die Zuordnung von Zahlen zu den Objekten ist eine Form der numerisch repräsentierten Rangfolge. Die bloße Rangfolge von zwei benachbarten Objekten mit den zwei ihnen zugeordneten Zahlen, z.B. $m(o_i) = 15$ und $m(o_j) = 30$, sagt aber nichts darüber aus, wie eng dieses Nachbarschaftsverhältnis ist, d.h. sie sagt nichts über die Größe der Abstände zwischen dem i-ten und dem j-ten Rang aus. Man kann nicht sagen, daß o_j doppelt so groß wie o_i ist oder doppelt so viel von dieser Eigenschaft besitzt.

2.3. Die Intervallskala

Gegenüber der Ordinalskala stellt die Intervallskala ein höheres Meßbarkeits- oder Quantifizierungsniveau dar, d.h. sie gibt eine größere Information über die Struktureigenschaften des Objektbereichs. Gewöhnlich betrachtet man eine Menge von Objekten, deren Eigenschaften eine Relationenstruktur aufweist, die die Anwendung einer Intervallskala gestattet, als quantitative Daten: m.a.W. erst Daten, die dieses oder ein höheres Meßbarkeitsniveau aufweisen, werden als quantitativ bezeichnet. Das ist aber eine Frage der Konvention. Das

höhere Skalenniveau kommt hier dadurch zum Ausdruck, daß neben der durch die Ordinalskala angegebene Rangfolge nunmehr auch die *Abstände* der einzelnen Größen bzw. ihrer Maßzahlen angegeben werden können. Wir können jetzt in einem bestimmten Sinne präzise Aussagen über die *Differenzen* in der Größe der einzelnen Objekte machen. Bei Erfassung durch eine Intervallskala können die ersten mathematischen Grundoperationen auf die Differenzen zwischen den verschiedenen Messungen angewendet und die Ergebnisse im Sinne von Aussagen über die zugrundegelegten Größen bzw. Eigenschaften der Objekte interpretiert werden.

Der Begriff des Abstands scheint uns zwar intuitiv geläufig zu sein, es müssen aber Bedingungen angegeben werden, die ihn genauer und vollständiger umreißen und ihn in den weiteren Zusammenhang der ersten mathematischen Grundoperationen, der Additivität, stellen. Wenn wir uns diese Grundoperation mit o darstellen, dann müssen die Größen des Objektbereichs folgende Bedingungen erfüllen, damit sie auf einer Intervallskala meßbar sind:

I_1 : Für alle $a, b \in M$ gilt : $a \ o \ b \sim x$ (Existenz)

I_2 : Für alle $a, b \in M$ gilt : $a \ o \ b \sim b \ o \ a \sim x$ (Kommutativität)

I_3 : Für alle $a, b, c \in M$ gilt : $a \ o \ (b \ o \ c) \sim (a \ o \ b) \ o \ c$ (Assoziativität)

I_4 : Für alle $a, b, c \in M$ gilt : $b \overset{<}{\sim} c \rightarrow (a \ o \ b) \overset{<}{\sim} (a \ o \ c)$
oder einfach $b \ R \ c \rightarrow (a \ o \ b) \ R \ (a \ o \ c)$ (Monotonie)

Sind diese Bedingungen erfüllt, so ist die Abbildung in die Menge der reellen Zahlen eindeutig bis auf lineare Transformationen. Der numerische Wert einer Messung auf einer Intervallskala kann in Meßwerte einer anderen Intervallskala durch eine Gleichung der Form $x' = \alpha x + \beta$ ($\alpha \neq 0$) transformiert werden. Als ein klassisches Beispiel einer Intervallskala kann man die Celsius-Temperaturskala ansehen. Wenn wir die Messungen $m(o_i) = 80$, $m(o_j) = 60$, $m(o_k) = 40$ haben, dann kann man sagen, daß die Differenz zwischen $m(o_i)$ und $m(o_j)$ gleich groß ist wie die Differenz zwischen $m(o_j)$ und

$m(o_k) : m(o_i) - m(o_j) = m(o_j) - m(o_k) = 20$. Darüber hinaus läßt sich noch folgendes sagen: eine Differenz $m(o_i) - m(o_k) = 40$ ist doppelt so groß wie zwischen $m(o_i)$ und $m(o_j)$. Dagegen kann man (bei dem Beispiel der Temperatur) nicht sagen, daß o_i doppelt so warm ist wie o_k . Auf diesem Niveau der Quantifizierbarkeit sind infolgedessen sowohl Differenzen und vielfache solcher Differenzen, durch die Additivität genau interpretierbar als auch der Nullpunkt einer Differenz.

Intervallskalen sind sehr oft das Beste, was man in der Arbeit auf den meisten wissenschaftlichen Gebieten überhaupt erreichen kann. Die Konstruktion einer solchen Skala erfordert meistens einen hohen Grad an schöpferischer Hypothesenbildung auf der einen Seite und eine gründliche Prüfung der Tragweite der dadurch erreichten Quantifizierung des Objektbereichs im Hinblick auf eine adäquate Repräsentation der zugrundeliegenden objektiven Struktur auf der anderen Seite. "One can do arithmetic to his heart's content on any set of numbers, but his results are not necessarily true statements about amounts of some property objects possess unless interval-scale requirements are met by the procedure for obtaining those numbers". ²⁵⁾

2.4. Die Ratio- oder Verhältnisskala

Die Ratioskala vereinigt alle Eigenschaften der übrigen Skalen in sich und hat als besonderes Merkmal das Vorhandensein eines *echten* oder *absoluten Nullpunkts*, d.h. eines Nullpunkts, der nicht willkürlich festgelegt ist, sondern Ergebnis der Meßoperationen ist, so daß beim Skalenwert Null die gemessene Eigenschaft oder Größe verschwindet, m.a.W. gleich Null ist. Das Vorhandensein eines absoluten Nullpunktes, der das zusätzliche Charakteristikum einer Verhältnisskala ist, darf nicht mit Produzierbarkeit, d.h. jeweiliger Erfassbarkeit dieses Punktes durch unmittelbare Meßoperationen, gleichgesetzt werden. Es gibt Verhältnisskalen, wie z.B. die der absoluten Temperatur (Kelvin), deren Nullpunkt niemals produziert werden kann. Bei dieser Skala kommt die zweite Gruppe mathematischer Grundoperationen, nämlich Multiplikation und Division, zur Anwendung.

Wenn wir die Messungen $m(o_i) = 80$ und $m(o_k) = 40$ haben (die z.B. das Gewicht oder die Länge von zwei Objekten o_i und o_k angeben), dann können wir das Verhältnis

$$\frac{m(o_i)}{m(o_k)} = 2$$

bilden und sagen, das Objekt o_i hat doppelt so viel von der gemessenen Eigenschaft (z.B. Gewicht oder Länge) wie das Objekt o_k .

Bei der Verhältnisskala braucht man nur die Maßeinheit festzulegen, um die Zuordnung von Zahlen zu Objekten eindeutig festlegen zu können.

Messungen auf einer Verhältnisskala sind indifferent gegenüber Ähnlichkeitstransformationen, die eine Einheitstransformation bedeuten gemäß der Gleichung

$$x' = \alpha x \quad (\alpha \neq 0)$$

Diese Transformationen wahren die Äquidistanz und den Nullpunkt und bedeuten rechnerisch die Multiplikation aller zu transformierenden Meßwerte mit einer Konstanten $\alpha \neq 0$.

3. Meßbarkeitsniveau und statistische Operationen

Nachdem wir einen Überblick über die Zuordnungsregeln und die mathematischen Eigenschaften der üblichen Skalenarten im vorhergehenden Abschnitt gegeben haben, können wir nun versuchen, einige Angaben zur Explizierung des in II. 1. angeführten dritten Aspekts zu machen, nämlich welche statistische Operationen den verschiedenen Skalenarten adäquat sind.

Die mathematische Operation, die bei Nominalskalen prinzipiell anwendbar ist, ist die des Zählens der Ereignisse oder Objekte in den einzelnen disjunkten Klassen. Somit ist der Modus auch auf dieser ersten Quantifizierungsstufe sinnvoll anwendbar; er gibt hier an, welche Klasse die meisten Objekte oder Ereignisse enthält, d.h. die "häufigste Klasse".

Ordinalskalen enthalten nicht nur eine Klassifikation des Objektbereichs, sondern geben darüberhinaus eine Ordnungsrelation zwischen den Klassen, m.a.W. die Richtung der Rangfolge an. Eine Maßzahl der zentralen Tendenz, die ordinal skalierte Daten adäquat beschreiben würde, müßte der Tatsache Rechnung tragen, daß die Klassen in aufsteigender oder absteigender Folge geordnet sind, was immer dies im konkreten Fall bedeuten mag. Der Median oder Zentralwert entspricht genau diesen Anforderungen, denn er ist der Punkt in der Skala, der die Verteilung der Ereignisse oder Objekte genau in der Mitte teilt: die eine Hälfte der Daten liegt unterhalb (in der gerichteten Folge) und die andere Hälfte oberhalb des Medians. Bei dieser Skalenart kann man auch die ersten Maßzahlen der Streuung anwenden, wie Spannweite und mittlerer Quartilabstand.

Intervallskalen geben sowohl die Richtung als auch die Abstände zwischen den einzelnen Klassen an, wobei diese Abstände in festen Einheiten gemessen werden. Die meisten der üblichen arithmetischen Operationen sind auf die Meßwerte dieser Skalen anwendbar. Die Berechnung des arithmetischen Mittels kann hier sinnvoll durchgeführt werden, denn das arithmetische Mittel ist der Punkt der Verteilung, von dem aus die Summe aller Differenzen der einzelnen Meßwerte Null ist; quantitative Aussagen über Differenzen (Abstände) zwischen Meßwerten einer Verteilung können also erst bei Intervallskalen gemacht werden. Entsprechendes gilt auch für die Streuungsmaße mittlere lineare Abweichung und Standardabweichung.

Schließlich wird die Berechnung des Variationskoeffizienten erst bei Verhältnisskalen angebracht sein, wenn ein echter Nullpunkt vorliegt. ²⁶⁾

4. Meßbarkeit und wissenschaftliche Argumentation

Solange man sich nur mit den formalen Kriterien der Meßbarkeit und den formalen Eigenschaften verschiedener Skalierungsarten beschäftigt, kann man auch keine Aussage darüber machen, ob der eine oder der andere, wie auch immer definierte Objektbereich, diesem oder jenem oder überhaupt einem Meßbarkeitsniveau zugänglich ist.

Eine a priori Entscheidung über die Anwendbarkeit einer Quantifizierungsstufe auf einen bestimmten Objektbereich ist nicht möglich.

Unter Quantifizierung wollen wir den Prozess der Aufdeckung und Beschreibung von Beziehungen zwischen den Elementen eines genau definierten Objektbereichs verstehen, und zwar solcherart, daß sie in formalen (meistens numerischen) relationalen Systemen abgebildet, d.h. durch sie repräsentiert werden können. Durch diese Umschreibung soll der Eindruck vermieden und der Versuchung schon im Ansatz vorgebeugt werden, als ob es sich bei der Quantifizierung um die ontologische Umwandlung von Qualität in Quantität, was natürlich absurd wäre, handelt. Denn was ontologisch qualitativ ist kann nicht durch "methodologische" Klimmzüge quantitativ gemacht werden.

"The type of scale achieved depends upon the character of the basic empirical operations performed. The operations are limited ordinarily by the nature of the thing being scaled and by our choice of procedures. . . ".²⁷⁾

Infolgedessen ist die begriffliche Festlegung einer Messungsskala oder konkreter eines Meßwertes nicht dadurch getan, daß man den einzelnen zu messenden Elementen des Objektbereichs Zahlen zuordnet, es müssen vielmehr die Operationen (Beobachtungen, Manipulationen, Kalkulationen) angegeben werden, die dem Zuordnungsprozess zugrundeliegen. Darüber hinaus, und da jede Messung die Konstruktion eines theoretischen Modells bedeutet, muß dieses Meßmodell der empirischen Prüfung an dem untersuchten Objektbereich bzw. an den zu messenden Elementen dieses Bereichs standhalten. Das soll gewährleisten, daß dem Objektbereich keine methodologisch erschaffene Faktizität aufgestülpt wird, die dann zusätzlich dadurch immunisiert wird, daß alle von dieser "Faktizität" abweichenden, mit den Forderungen des gewählten Maßbarkeitsniveau unvereinbare Daten als "Fehlergrößen" deklariert werden.²⁸⁾

"Oft will man aber auch einen Begriff messen, dessen Inhalt durch die Operationsangaben nicht erschöpfend gekennzeichnet ist; z.B. wäre dann die Reaktion einer Person auf bestimmte Fragen nur ein

Indiz für ihre “politische Einstellung”. Man hat dann die “Einstellung” unabhängig von den Operationen charakterisiert und damit gewissermaßen hypostasiert. Das ist sinnvoll, wenn dieser Begriff noch in anderen Zusammenhängen als in denjenigen, mit denen man gerade operiert, vorkommt, und wenn er in diesen anderen Zusammenhängen auch durch andere operationale Indizien beobachtbar gemacht wird. In der Regel handelt es sich dann um eine sogenannte hypothetische Entität im Rahmen einer umfassenden Theorie, aus welcher man vielerlei Hypothesen ableiten kann; als Beispiel diene der physikalische Begriff des “Elektron”, vor allem aber viele Begriffe der Psychologie, etwa bestimmte “Motive”, “Triebe” u.ä.m.” ²⁹⁾

Ähnlich verhält es sich mit der von Osgood und seinen Mitarbeitern entwickelte Methode zur Messung von Bedeutung durch das sogenannte “semantische Differential”. ³⁰⁾

“Das semantische Differential ist im wesentlichen eine Kombination von kontrollierter Assoziation und Skalierungsverfahren. Wir geben der Versuchsperson einen zu differenzierenden Begriff und eine Reihe bipolarer adjektivischer Skalen, anhand derer die Differenzierung vorzunehmen ist, wobei es die einzige Aufgabe der Versuchsperson ist, für jedes Einzelelement (Verbindung eines Begriffs mit einer Skala) die Richtung ihrer Assoziation und deren Intensität auf einer Sieben-Stufen-Skala anzugeben. Die Schwierigkeit der Methode liegt natürlich in der Auswahl des Samples deskriptiver bipolarer Termini”. ³¹⁾

Wenn wir zwei Wörter nehmen “Baum” und “Stein”, und Testpersonen fragen, unter welchen Aspekten sie diese zwei Ausdrücke unterscheiden, dann werden wir unter anderen beispielsweise auch diese Antworten erhalten: Ein Baum ist *lebendig* (belebt), während ein Stein *tot* (unbelebt) ist; oder ein Baum ist relativ *flexibel*, während ein Stein *starr*, (*unbiegsam*) ist. Entsprechend werden sie ihre Eintragungen auf die jeweiligen bipolaren adjektivischen Skalen (mit sieben Einheiten) eintragen.

Es ist klar, daß zu einer umfassenden Bestimmung der Bedeutung von “Baum” und “Stein”, d.h. wie diese Wörter von jenen, die diese

Sprache benutzen, verstanden und gebraucht werden, eine große Anzahl von solchen bipolaren Termini erforderlich sein wird. Bei seinen Untersuchungen kam Osgood zu einer Ansammlung von 50 verschiedenen solchen Paaren, die er unter Verwendung der Faktorenanalyse zu drei Gruppen, Grunddimensionen oder Faktoren zusammengefaßt hat:

- 1) Bewertungsfaktor
- 2) Potenzfaktor
- 3) Aktivitätsfaktor.

Aus diesen drei bipolaren Grunddimensionen läßt sich nunmehr ein semantischer Raum konstruieren. In diesem Raum werden die jeweiligen Testergebnisse eingetragen. Nun formuliert Carroll: "For some experimental purposes, it is useful to measure (1) the *polarity* of concepts, that is, their distance from the center or origin in the semantic space, and (2) the *distances* between concepts, that is, how different their "meanings" are". ³²⁾

Unabhängig von anderen Einwänden, die gegen das Verfahren des "semantischen Differentials" und seiner Leistungsfähigkeit und Adäquatheit erhoben werden können, tritt hier ein rein technischer hinzu. Durch die oben beschriebene Versuchsanordnung wird offensichtlich eine Ordinalskala konstruiert, zwar nicht eine, die der hier beschriebenen voll entspricht, sondern eine, die das Niveau der Intervallskalen noch nicht erreicht hat. Infolgedessen ist eine Aussage über die Abstände der verschiedenen Punkte des semantischen Raumes nicht statthaft. Darüber hinaus wäre auch unter der Voraussetzung, daß Aussagen über Differenzen oder Abstände der semantischen Punkte möglich wären, eine Schlußfolgerung über die Frage, wie verschieden die Eigenschaften sind, die sie repräsentieren, nämlich "Bedeutung", immer noch nicht statthaft:

Das Problem des Meßbarkeitsniveaus ist kein Problem der Mathematik oder der Statistik. Denn die Statistik befaßt sich mit numerischen Daten, und wenn man auch durch die statistischen Methoden Schlußfolgerungen unter Zugrundelegung dieser numerischen Daten ziehen kann, es gibt keine Möglichkeit von der Statistik her zu zeigen, daß

die numerischen Daten tatsächlich keine sagen wir intervallskalierten Messungen von irgendwelchen Objekten oder Eigenschaften sind. ³³⁾

Denn die Anwendung irgendwelcher statistisch-arithmetischer Operationen auf numerische Daten ergibt numerische Resultate, die formal richtig sind. "Even if the numbers are the purest nonsense, having no relation to real magnitudes or properties of real things, the answers are still right *as numbers*. The difficulty comes with the interpretation of these numbers back into statements about the real world. If nonsense is put into the mathematical system, nonsense is sure to come out". ³⁴⁾

Wenn wir z.B. einen Sprachtest zur Prüfung der Fähigkeiten von Studenten in ihrer Muttersprache durchführen und dabei feststellen, daß ein Student keine einzige Testfrage korrekt beantwortet hat, dann können wir natürlich feststellen, wenn wir einfach die korrekten Antworten oder anders ausgedrückt die Testerfolge zählen, daß in diesem Fall tatsächlich ein absoluter oder echter Nullpunkt vorliegt. Ein Testausgang von "null korrekten Antworten" bedeutet nicht "Null" - Fähigkeit in dieser Sprache. Das heißt aber, daß die Skalierung von "Fähigkeit in der Muttersprache" keinen absoluten Nullpunkt enthält. Infolgedessen kann auch nicht gesagt werden, daß ein Testergebnis von 50 korrekten Antworten eine doppelt so große Fähigkeit in der Sprache indiziert wie ein solches von 25 korrekten Antworten. Die Intervallskala erscheint daher als das adäquate Meßbarkeitsniveau für den Untersuchungsgegenstand "Fähigkeit in der Sprache" (natürlich solange keine weiteren Einwände nunmehr gegen dieses Niveau erhoben werden). ³⁵⁾

Es folgt daher, daß die Anwendbarkeit der verschiedenen Meßbarkeitsstufen nicht a priori behauptet werden kann, sondern empirisch nachgewiesen werden muß. Denn jede Messung bedeutet die Konstruktion eines theoretischen Modells, das empirisch überprüft werden muß.

"Vom Gesichtspunkt der Methodologie", habe ich an anderer Stelle ausgeführt ³⁶⁾, "erfüllen Messungen zwei wichtige Funktionen.

Sie bilden einerseits eine zuverlässige Grundlage zur Schlichtung von Kontroversen über die Tragweite und Wichtigkeit von Existenzaussagen, und zwar wegen der durch die grundsätzliche Wiederholbarkeit der Meßoperationen gewährleisteten intersubjektiven Erfahrung. Andererseits sind gemessene Daten durch den operationalen Rahmen genau definiert, so daß durch die Meßnormen feinere Differenzierungen und daher präzisere Beschreibungen als durch die Umgangssprache erreicht werden ” “Sobald man aber diese Stufe abstrakter Überlegungen verläßt und sich dem realen Objektbereich zuwendet, wird man nicht nur mit den bekannten Problemen der Objektidentifikation, sondern auch mit den speziellen Dimensionen der sozialwissenschaftlichen Erkenntnisproblematik konfrontiert, die sich epigrammatisch mit der Umformung kommunikativer Erfahrung und intensionalen Handelns in Daten ausdrücken läßt”.³⁷⁾

Meßbarkeitsniveaus und zulässige statistische Maßzahlen

Skala	empirische Grundoperationen	Charakterisierung der mathematischen Struktur	Maßzahlen der zentralen Tendenz	Maßzahlen der Streuung
Nominalskala	Bestimmung von Gleichheit	Permutations-Transformation	Modus	Information H
Ordinalskala	Bestimmung von "größer" oder "kleiner"	monotone Transformation $x' = f(x)$, wobei $f(x)$ eine monoton wachsende Funktion	Median	Perzentile
Intervallskala	Bestimmung von Gleichheit oder Differenz von Intervallen	lineare Transformation $x' = \alpha x + \beta$ ($\alpha > 0$)	arithmetisches Mittel	mittlere lineare und Standard-Abweichung
Verhältnisskala	Bestimmung von Quotientengleichheit	multiplikative Transformation (Ähnlichkeits- oder Einheiten-Transformation) $x' = \alpha x$, ($\alpha > 0$)	geometrisches und harmonisches Mittel	

D. Wahrscheinlichkeitsrechnung.

1. Kausalität und Zufall.

Mit Hilfe der Wahrscheinlichkeitsrechnung werden in den verschiedenen Wissenschaften in der theoretischen Sphäre als auch in der praktischen Anwendung *zufällige* Erscheinungen untersucht und über die Gesetzmäßigkeiten, nach denen die untersuchten Erscheinungen verlaufen, Aussagen gemacht.

Demgegenüber gibt es in unserer Umwelt (Gesellschaft, Natur, Technik) Vorgänge, die unter einem genau bekannten und definierten Bedingungskomplex mit Bestimmtheit eintreten. Solche determinierte Erscheinungen treten häufig in der klassischen Physik auf. Durch das Prinzip der Kausalität wird einem gewissen Bedingungskomplex eine eindeutige Folge zugeordnet. Aus diesem Prinzip ergeben sich zwei mögliche Fälle: wenn bei jeder Realisierung eines bestimmten Bedingungskomplexes ein Ereignis eintritt, so nennen wir es ein *sicheres Ereignis*, kann es andererseits unter dieser Bedingungskonstellatation auf keinen Fall eintreten, so nennen wir es ein *unmögliches Ereignis*. Die Ursache B hat die eindeutige Wirkung A:

$$B \longrightarrow A$$

“Ein Ereignis ist zufällig, wenn es nicht sicher und nicht unmöglich ist”³⁸⁾. Das soll aber nicht besagen, daß der Realisationsablauf nicht kausal bestimmt wäre, sondern nur, daß wir das Ergebnis einer konkreten Realisation nicht im voraus bestimmen können. Das eingetretene Ereignis ist das Ergebnis einer kausalen Abfolge, nur wissen wir nicht im voraus, welches der unter dieser Bedingungskonstellatation möglichen Ereignisse tatsächlich eintreten wird.

Ein bestimmter Bedingungskomplex ist notwendig und hinreichend für das Eintreten des sicheren Ereignisses, d.h. dieses einzelne Ereignis folgt mit Sicherheit aus dem Bedingungskomplex, während bei zufälligen Ereignissen die Bedingungskonstellatation nur für eine Ereignismenge notwendig und hinreichend ist.

$$B \longrightarrow E = \{A_1, A_2, \dots, A_n\}$$

“Der Laie hat oft die Vorstellung, daß überall dort, wo der Zufall regiert, die Wissenschaft ein Ende findet. Aber diese Meinung ist falsch.

Der Zufall ist meßbar! Das Maß für die Neigung oder Leichtigkeit des Eintretens eines zufälligen Ereignisses ist seine Wahrscheinlichkeit”³⁹⁾

2. Ereignisraum und Elementarereignis.

Alle zufälligen Ereignisse, die für eine durch einen klar definierten Bedingungskomplex bestimmte Problemstellung eintreten können, bilden eine Menge von Elementarereignissen, die man den Ereignisraum E bezeichnet. Mathematisch besteht also ein Ereignisraum aus einer Menge. Zwischen der Menge E der Elementarereignisse und der Menge M der Merkmalsmodalitäten besteht eine unmittelbare Entsprechung. Die Menge E ist die Realisationsstufe der Menge M , d.h. ein Elementarereignis ist die Realisation einer Merkmalsmodalität. Im Weiteren werden wir E und M als identisch betrachten. Durch die mathematische Festlegung des Ereignisraumes ergibt sich die mathematische Definition eines beliebigen Ereignisses als Teilmenge der Menge E :

$$A_i \subseteq E$$

Bei der Darstellung der Beziehungen zwischen zufälligen Ereignissen kann man sich der mengentheoretischen Schreibweise bedienen. Zur Veranschaulichung dieser Beziehungen können wir die Wurfresultate mit einem idealen sechsseitigen Würfel heranziehen.

- 1) $A_1 \subset A_2$: wenn das Ereignis A_1 eintritt, so ist das Ereignis A_2 auch eingetreten.
Beispiel: Wenn $A_1 = \{1\}$ und $A_2 = \{1, 3, 5\}$, dann gilt $A_1 \subset A_2$

- 2) $A_1 \cup A_2 = A_3$: Die logische Summe der Ereignisse A_1 und A_2 tritt ein, wenn mindestens eines der Ereignisse A_1 oder A_2 eingetreten ist.
 Beispiel: Wenn $A_1 = \{4\}$ und $A_2 = \{1,3,5\}$ ist, dann tritt A_3 genau dann ein, wenn entweder 1 oder 3 oder 4 oder 5 geworfen wird; $A_3 = \{1,3,4,5\}$.
- 3) $A_1 \cap A_2 = A_3$: Das logische Produkt von A_1 und A_2 tritt ein, wenn sowohl A_1 als auch A_2 eingetreten ist.
 Beispiel: Wenn $A_1 = \{3\}$ und $A_2 = \{1,3,5\}$, dann ist $A_3 = \{3\}$.
- 4) $A \cup A' = E$: Die logische Summe des Ereignisses A und seines komplementären Ereignisses A' ist das sichere Ereignis E .
 Beispiel: Wenn $A = \{2\}$ ist $A' = \{1,3,4,5,6\}$ und nach 2) ist ihre Vereinigung $E = \{1,2,3,4,5,6\}$, was das sichere Ereignis ist.
- 5) Das unmögliche Ereignis wird durch das Symbol ϕ ausgedrückt.
 Beispiel: $\{8\} = \phi$
- 6) $A_1 \cap A_2 = \phi$: Die Ereignisse A_1 und A_2 schließen sich gegenseitig aus, d.h. ihr gleichzeitiges Eintreten ist unmöglich. Solche Ereignisse heißen disjunkt (oder unvereinbar).
 Beispiel: Wenn $A_1 = \{2\}$ und $A_2 = \{1,3,5\}$, dann ist $A_1 \cap A_2 = \phi$.
 Es gilt weiterhin $A \cap A' = \phi$.
 Aus 4) und 5) ergibt sich, daß E und ϕ komplementäre Ereignisse sind.

Außer den einelementigen Mengen, die wir Elementarereignisse genannt haben, können wir auch zusammengesetzte Ereignisse bilden.

Beim Würfelspiel würde das bedeuten, daß wir außer den Elementarereignissen $1, 2, \dots, 6$ auch Ereignisse betrachten, wie das Ereignis $\{2,5\} \subset \{1,2, \dots, 6\}$, d.h. entweder eine 2 oder eine 5 zu würfeln, oder das Ereignis $\{2,4,6\} \subset \{1,2, \dots, 6\}$, d.h. eine gerade Zahl zu würfeln usw. Diese Überlegungen führen zu der Konstruktion einer Menge, die alle möglichen Realisationen vom Merkmalsmodalitäten als Untermengen hat, d.h. sowohl die Elementarereignisse als auch die zusammengesetzten Ereignisse als auch das sichere und das unmögliche Ereignis. Diese Menge wird Potenzmenge genannt, $P^*(E)$.

Die Anzahl der Untermengen der Potenzmenge können wir durch folgende Überlegungen ermitteln:

Wenn der Ereignisraum leer ist ($E = \emptyset$), dann enthält die Potenzmenge ein Element, nämlich die Nullmenge. Es ist $P^*(E) = \{\emptyset\}$.

Wenn $E = \{0,1\}$, dann ist $P^*(E) = \{\{0,1\}, \{0\}, \{1\}, \emptyset\}$, d.h. sie enthält vier Elemente.

Wenn $E = \{1,2,3\}$, dann enthält die Potenzmenge acht Elemente, nämlich $P^*(E) = \{\{1,2,3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1\}, \{2\}, \{3\}, \emptyset\}$.

Daraus kann man schon die generelle Formel für die Anzahl der Elemente einer Potenzmenge ermitteln: sie lautet 2^n , wobei n die Anzahl der Elementarereignisse ist.

Die Gesamtheit aller zufälligen Ereignisse bilden eine Potenzmenge von E , $P_F^*(E)$, die Ereignisfeld genannt wird, mit folgenden Eigenschaften:

- (1) $E \in P_F^*(E)$ und
- (2) wenn $A_1, A_2 \in P_F^*(E)$, dann enthält das Ereignisfeld auch alle zufälligen Ereignisse, die durch die binären Operationen 2), 3), 4) entstehen d.h.
 $A_1 \cap A_2 \in P_F^*(E)$.

3. Begriff der Wahrscheinlichkeit.

Der Begriff der Wahrscheinlichkeit taucht sehr oft in der Alltagssprache auf. Dieser Begriff kann aber wegen seines unpräzisen Inhalts nicht als Grundlage von wissenschaftlichen Theorien herangezogen werden. Eine präzise Definition wäre deshalb notwendig. Trotz der großen Verbreitung der Wahrscheinlichkeitsrechnung in den verschiedenen Wissenschaftszweigen, die sich auf mathematisches Rasonieren (ganz oder teilweise) stützen, stößt die Interpretation des Begriffs auf große Schwierigkeiten.

Aus der langen Diskussion über den Wahrscheinlichkeitsbegriff lassen sich drei verschiedene Schulen erkennen, die folgende Typen der Wahrscheinlichkeit unterscheiden:

(1) objektive a priori Wahrscheinlichkeit. Diesem klassischen Wahrscheinlichkeitsbegriff liegt folgende Vorstellung zugrunde:

Gegeben ist ein Ereignisfeld, das insgesamt n Elementarereignisse A_1, A_2, \dots, A_n enthält mit den Eigenschaften

$$\bigcup_{i=1}^n A_i = E \quad \text{und}$$

$$A_i \cap A_j = \phi, \text{ wobei } i \neq j \text{ (} i, j = 1, 2, \dots, n \text{)}.$$

Man betrachtet die einzelnen Elementarmodalitäten als gleichwahrscheinlich oder gleichmöglich. Dabei geht man vom Prinzip der Indifferenz (oder Prinzip vom unzuzeichnenden Grund) aus. Es besagt: wenn kein a priori Grund besteht, die Chance für das Auftreten einer Elementarmodalität für größer zu halten als die für das Auftreten einer anderen, dann betrachtet man alle Elementarereignisse als gleich möglich

$$P(A_1) = P(A_2) = \dots = P(A_n) = \frac{1}{n} = \text{const.}$$

Wenn jetzt ein beliebiges Ereignis B gegeben ist, dann kann es in einer

endlichen Anzahl m von Elementarereignissen A_1, A_2, \dots, A_m ,
 $0 \leq m \leq n$, zerlegt werden, und seine Wahrscheinlichkeit wird
 durch

$$P(B) = \frac{m}{n} \quad \text{oder}$$

$$P(B) = \frac{\text{für } B \text{ günstige Elementarereignisse}}{\text{alle mögliche Elementarereignisse}}$$

gegeben.

Für $m = 0$ ist $B = \emptyset$ und $P(B) = 0$,

für $m = n$ ist $B = E$ und $P(B) = 1$

Klassische Definition (stammt von P. - S. de Laplace):

Wenn ein Merkmal in n verschiedenen, sich gegenseitig ausschließen-
 den Modalitäten (Elementarereignissen) eintreten kann, die alle gleich-
 wahrscheinlich sind, und wenn m von ihnen als Erfolg angesehen wer-
 den, dann ist die Wahrscheinlichkeit eines Erfolges bei einem einzelnen
 Zufallsexperiment

$$P(A_i) = \frac{m}{n}$$

Beispiel

Wie groß ist die Wahrscheinlichkeit, bei einem Wurf mit einem regel-
 mäßigen Würfel eine Sechs zu werfen?

— Als Erfolg oder als günstiges Ereignis wird die *Zahl* 6 angesprochen,
 d.h. ein günstiges Ereignis aus 6 möglichen.

Die Wahrscheinlichkeit ist also

$$P(A) = \frac{\text{günstige Ereignisse}}{\text{mögliche Ereignisse}} = \frac{1}{6}$$

Zwei wichtige Eigenschaften charakterisieren, unter anderem, diesen
 Wahrscheinlichkeitsbegriff.

- 1) Er setzt Symmetrie der Ereignisse voraus.
- 2) Seine Ableitung erfolgt durch abstraktes Rasonieren und ist unabhängig von der Erfahrung (Experiment).
(beruht auf Merkmalen).

Der klassische Wahrscheinlichkeitsbegriff ist nur auf Ereignisfelder mit einer endlichen Anzahl von Elementarereignissen anwendbar. Die Berechnung der Wahrscheinlichkeit eines zufälligen Ereignisses ist nach dieser Definition oft nicht möglich, da nicht immer die Möglichkeit besteht, alle gleichwahrscheinlichen Elementarereignisse zu ermitteln.

- (2) Objektive a posteriori Wahrscheinlichkeit, auch Häufigkeitswahrscheinlichkeit genannt. Wir können zwei verschiedene Richtungen unterscheiden:

- (2.1) Die erste Richtung geht von der relativen Häufigkeit eines Ereignisses aus und definiert die Wahrscheinlichkeit folgendermaßen:

Wenn in einer Reihe von n Versuchen ein Ereignis A k -mal eingetreten ist, so *ist* die Wahrscheinlichkeit für das Auftreten von A in einem einzelnen Versuch

$$P(A) = \frac{k}{n}$$

Diese Definition zeichnet sich dadurch aus, daß wenn der Versuch sehr oft (n) wiederholt wird, und dabei das Ereignis A k -mal eingetreten ist, dann wird die relative Häufigkeit k/n praktisch (beinah, approximativ) der wahren Wahrscheinlichkeit P gleich sein. Charakteristisch für diese Definition ist, daß

- a) sie eine große Anzahl von Versuchen voraussetzt,
- b) sie statistische Regularität voraussetzt,
- c) die Wahrscheinlichkeit $P(A)$ durch die relative Häufigkeit von A gemessen wird,
- d) sie von der Erfahrung (Experiment) ausgeht.

In der Formulierung von Kolmogoroff heißt es:

“Man kann praktisch sicher sein, daß, wenn man den Komplex der Bedingungen S eine große Anzahl von n Malen wiederholt und dabei durch m die Anzahl der Fälle bezeichnet, bei denen das Ereignis A stattgefunden hat, das Verhältnis von m/n sich von $P(A)$ nur wenig unterscheidet.”⁴⁰⁾

- (2.2) Die zweite Richtung geht zwar auch von der relativen Häufigkeit der elementaren Merkmalsmodalität aus, sie definiert aber die Wahrscheinlichkeit als *den Grenzwert* der relativen Häufigkeit k/n wenn $n \rightarrow \infty$ strebt: d.h.

$$P(A) = \lim_{n \rightarrow \infty} \frac{k}{n}$$

R.v. Mises ist der Hauptvertreter dieser Richtung.

- (3) Subjektive Wahrscheinlichkeit. Die subjektivistische Auffassung charakterisiert Savage so: “Personalistic views hold that probability measures the confidence that a particular individual has in the truth of a particular proposition, for example the proposition that it will rain tomorrow. These views postulate that the individual concerned is in someways ‘reasonable’, but they do not deny the possibility that two reasonable individuals faced with the same evidence may have different degrees of confidence in the truth of the same proposition”.⁴¹⁾

Wie diese Ausführungen zeigen, wird die Wahrscheinlichkeit der Realisierung einer Merkmalsmodalität A (des Auftretens des Ereignisses A) als Grad des rationalen Glaubens (rational oder dispositional belief), des Vertrauens (confidence) eines Individuums an diese bestimmte Aussage, oder anders ausgedrückt, an das Eintreten dieses Ereignisses betrachtet.

Auf die umfangreiche Axiomatik der subjektiven Wahrscheinlichkeit wird hier nicht weiter eingegangen.

4. Die Axiomatisierung der Wahrscheinlichkeitsrechnung

Die als statistisch bezeichnete Definition des Wahrscheinlichkeitsbegriffs, wonach die Wahrscheinlichkeit aus der relativen Häufigkeit des Auftretens eines zufälligen Ereignisses abgeleitet wird, ist für die überwiegende Anzahl der praktischen Anwendungssituationen in den verschiedenen wissenschaftlichen Disziplinen ausreichend und befriedigend. Für den mathematisch-theoretischen Aufbau der Wahrscheinlichkeitsrechnung ist aber eine exakte Systematisierung und eine widerspruchsfreie Begriffsbildung unerlässlich. Ableitbarkeit mit Hilfe eines mathematisch-logischen Kalküls setzt eine erste Quelle von anerkannten Grundsätzen (Grundeigenschaften) voraus, d.h. der Aufbau eines theoretischen Komplexes muß auf ein System von bereits bekannten Sätzen zurückzuführen sein - und durch logische Operationen von ihm ableitbar sein. Dieser Regress auf solche bekannte Sätze kann natürlich nicht ad infinitum geführt werden. Solche in den verschiedenen Wissenschaften konventionell festgesetzte und logisch nicht mehr ableitbare Sätze nennt man Axiome. Ihre Gültigkeit ist selbstevident, d.h. sie beruht auf Erfahrung. Durch die widerspruchsfreie Zusammenführung einer Anzahl von Axiomen entsteht ein Axiomensystem.

Das von Kolmogoroff eingeführte Axiomensystem umfaßt folgende drei Axiome:

1. *Axiom.* Jedem zufälligen Ereignis A ($A \in \mathcal{P}_F(E)$)

wird eine reelle Zahl $P(A)$, die Wahrscheinlichkeit von A , zugeordnet, die folgende Bedingung erfüllt:

$$0 \leq P(A) \leq 1$$

2. *Axiom.* Die Wahrscheinlichkeit des sicheren Ereignisses ist gleich 1 :

$$P(E) = 1$$

Die Aussage dieses Axioms ist nicht unbedingt umkehrbar, d.h. bei einem Ereignis mit einer Wahrscheinlichkeit gleich 1 können wir nicht in jedem Fall den Schluß ziehen, daß wir es mit einem sicheren Ereignis zu tun haben. Das gleiche gilt für die Beziehung zwischen $P(A) = 0$

und $A = \phi$: die Aussagen sind auch hier nicht unbedingt umkehrbar.

Die Umkehrbarkeit dieser Beziehungen ist bei diskreten Zufallsvariablen aus einem endlichen Ereignisraum gegeben. Bei stetigen Zufallsvariablen aus unendlichen Ereignisräumen ist die Umkehrbarkeit nur tendenziell gegeben.

3. *Axiom.* Sind die zufälligen Ereignisse A_1 und A_2 disjunkt, so ist die Wahrscheinlichkeit für das Eintreten entweder des Ereignisses A_1 oder des Ereignisses A_2 gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse, d.h. wenn $A_1 \cap A_2 = \phi$, dann ist

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

* Anders ausgedrückt: Die Wahrscheinlichkeit der Vereinigung disjunkter zufälliger Ereignisse ist die Summe ihrer einzelnen Wahrscheinlichkeiten.

Erweitertes 3. Axiom. Sind die zufälligen Ereignisse A_1, A_2, \dots, A_n paarweise disjunkt, dann ist

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

$$= \sum_{i=1}^n P(A_i).$$

Diese Erweiterung zeigt, daß die Potenzmenge einer Menge E von Elementarereignissen auch ein mathematischer Mengenkörper ist, d.h.

$$o(A_1, A_2, \dots, A_n) \in P^*_F(E) = K(E).$$

Unter dem Mengenkörper über E verstehen wir also eine Menge $K(E)$, welche Vereinigung, Durchschnitt und Komplementärmenge der Teilmengen ebenfalls als Elemente enthält.

Aus diesen Axiomen wollen wir hier einige Sätze, die für die Wahrscheinlichkeitsrechnung wichtig sind, exemplarisch vorführen, ohne

auf die Beweisführung einzugehen, die leicht zu liefern ist:

1. Satz. $P(A \cup A') = P(A) + P(A') = 1$

2. Satz. Wenn $A_1 \subset A_2$, dann ist $P(A_1) \leq P(A_2)$

3. Satz. Für zwei beliebige Ereignisse A_1 und A_2 ,

die sich nicht unbedingt ausschließen, gilt, daß die Wahrscheinlichkeit für das Eintreten von A_1 oder von A_2 gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse weniger die Wahrscheinlichkeit für das Eintreten von A_1 und A_2 ist.

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

Wenn die Ereignisse disjunkt sind, d.h. $A_1 \cap A_2 = \emptyset$, dann ergibt sich das dritte Axiom.

Das Tripel $(E, K(E), P)$, d.h. das Tripel aus der Menge der Elementarereignisse, dem Mengenkörper über E und der jedem zufälligen Ereignis zugeordneten Wahrscheinlichkeit, bildet einen Wahrscheinlichkeitsraum. Für ihn gelten die für seine drei Komponenten schon ausgeführten Eigenschaften.

5. Zusammengesetzte Ereignisse

Wir haben weiter oben ausgeführt, daß durch Operationen unter den Elementarereignissen eines Ereignisraumes zusammengesetzte ⁴²⁾ Ereignisse gebildet werden, und gelangten dadurch zu einer Erweiterung des Ereignisbegriffs, die zum Mengenkörper über E (Eignisraum) führten. Wir können nunmehr einige Beobachtungen über diese Ereignisse machen.

Wenn wir das logische Produkt zweier Ereignisse bilden $A \cap B$, dann tritt dieses Ereignis dann ein, wenn irgendein Elementarereignis, das zu dieser Menge gehört eintritt. Ein Element von $A \cap B$ muß sowohl ein Element von A als auch von B sein, d.h. $A \cap B \Leftrightarrow c \in A \wedge c \in B$ und infolgedessen treten sowohl A als auch B ein, wenn $A \cap B$ ein-

tritt. Andererseits ist ein Element von $A \cap B$ auch ein Element von $A \cup B$, d.h. $c \in A \cap B \rightarrow c \in A \cup B$; infolgedessen wird bei der Realisierung von $A \cap B$ auch $A \cup B$ realisiert.

Die Wahrscheinlichkeit errechnet sich in der üblichen Art

$$P(A \cap B \cap C) = \frac{\text{Anzahl der Elementarereignisse in } A \cap B \cap C}{\text{Gesamtanzahl der Elementarereignisse.}}$$

Die Wahrscheinlichkeit der logischen Summe der Ereignisse A, B, C , d.h. für das Eintreten von entweder A oder B oder C oder A, B, C , wird errechnet durch

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

und für disjunkte Ereignisse

$$P(A \cup B \cup C) = P(A) + P(B) + P(C).$$

6. Bedingte Wahrscheinlichkeit und Unabhängigkeit

Wenn wir zwei Teilmengen A_1 und A_2 aus dem Ereignis = Mengenkörper $K(E)$ nehmen und eine davon als Bezugsmenge nehmen (hier A_2), mit $P(A_2) > 0$, dann bezeichnet man als bedingte Wahrscheinlichkeit von A_1 gegeben A_2 (oder unter der Bedingung A_2 , oder nachdem A_2 eingetreten ist) das Verhältnis der Wahrscheinlichkeit des Durchschnitts $A_1 \cap A_2$ zur Wahrscheinlichkeit der Bezugsmenge A_2 .

$$P(A_1 | A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} \quad \text{wobei } P(A_2) > 0$$

Es gilt auch:

$$P(A_2 | A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} \quad ; \quad P(A_1) > 0$$

oder in Worten:

$$P(A_1 | A_2) = \frac{\text{Anzahl der Elementarmodalitäten in } A_1 \cap A_2}{\text{Anzahl der Elementarmodalitäten in } A_2}$$

Durch die bedingte Wahrscheinlichkeit wird ein beliebiges Ereignis A_1 nicht mehr auf den ursprünglichen Bedingungskomplex bezogen, sondern auf einen neuen, den wir durch Hinzuziehung der Bedingung A_2 aus der ursprünglichen Versuchsanordnung gewonnen haben. Durch diese Änderung des ursprünglichen Bedingungskomplexes restringieren wir die Beobachtung auf eine Untermenge von Ereignissen. Das Verhältnis der Wahrscheinlichkeiten, die wir den Ereignissen dieser Untermenge zuordnen, ist das gleiche wie in der ursprünglichen Menge, ihre Summierung ergibt aber hier eins. Dieser Sachverhalt läßt sich für Massenerscheinungen folgendermaßen formulieren: (Siehe auch in diesem Kapitel, Absatz 2.1.)

Wenn in einer Reihe von n Versuchen ein Ereignis A_2 m -mal eingetreten ist und bei diesen m Versuchen k -mal ($k \leq m$) das Ereignis A_1 zusammen mit A_2 realisiert wurde, dann bezeichnet man den Quotienten $\frac{k}{m}$ bedingte Wahrscheinlichkeit von A_1 unter der Bedingung A_2 .

$$P(A_1 | A_2) = \frac{k}{m}$$

Beispiel: Eine Münze wird dreimal geworfen. Wir bezeichnen A = Adler und K = Kopf.

Dann ist

$$E = \{(K, K, K), (K, K, A), (K, A, K), (K, A, A), (A, K, K), (A, K, A), (A, A, K), (A, A, A)\}$$

Wenn wir jetzt zwei Ereignisse folgendermaßen definieren:

$$B = \{(K, K, K), (K, K, A), (K, A, K), (A, K, K), (A, K, A), (A, A, K)\}$$

und

$$C = \{(K, A, A), (A, K, A), (A, A, K)\}$$

Das Ereignis B ist so zu interpretieren, daß bei dem zweiten und dritten Wurf mindestens einmal Kopf geworfen wird, und C ist das Ereignis, daß von den drei Würfeln zwei immer Adlerwürfe sind. Die Wahrscheinlichkeit des Eintretens der einzelnen Ereignisse aus dem Ereignisraum E ist $P(m_i) = 1/8$. Wir können nun die Wahrscheinlichkeit der Ereignisse B und C ermitteln.

$$P(B) = 6/8 = 3/4 \text{ und } P(C) = 3/8$$

Der Durchschnitt der Ereignisse B und C ist

$$B \cap C = \{(A, K, A), (A, A, K)\} \quad \text{und die Wahrscheinlichkeit des Durchschnitts,}$$

$$P(B \cap C) = 2/8 = 1/4$$

Die bedingten Wahrscheinlichkeiten lassen sich nun leicht ermitteln

$$P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3} \quad \text{und}$$

$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{1/4}{3/8} = \frac{2}{3}$$

Die Interpretation der bedingten Ereignisse und ihrer Wahrscheinlichkeit läßt sich leicht aus der Definition der Ereignisse B und C ableiten.

Aus diesen Überlegungen zur Wahrscheinlichkeit von bedingten Ereignissen läßt sich der Multiplikationssatz der Wahrscheinlichkeitsrechnung ableiten.

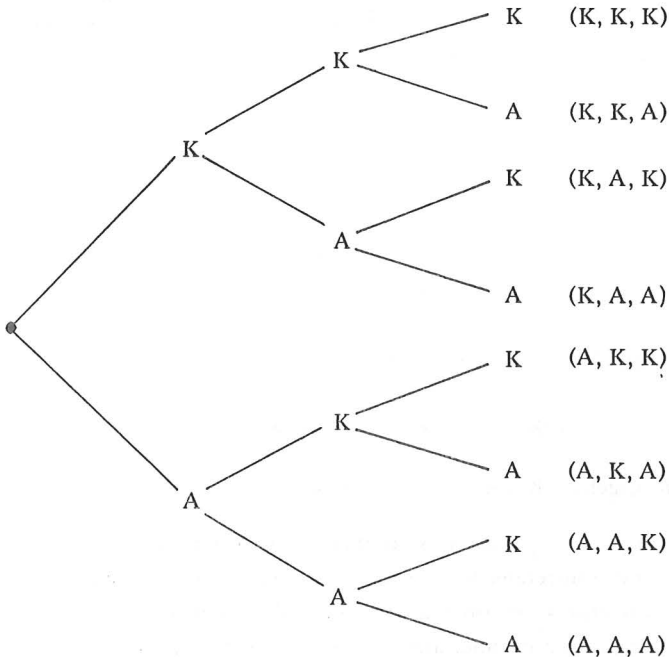
Wenn wir die Formel der bedingten Wahrscheinlichkeit nach $P(A_1 \cap A_2)$ auflösen, erhalten wir den Multiplikationssatz von beliebigen Ereignissen:

$$4. \text{ Satz. } P(A_1 \cap A_2) = P(A_2) \cdot P(A_1|A_2).$$

Das ist der "sowohl-als auch" Satz der Wahrscheinlichkeitsrechnung.

Bevor wir zum zweiten Teil dieses Abschnitts übergehen, wollen wir hier eine Darstellung des Ereignisraumes angeben, die in der Linguistik zur Darstellung eines anderen Sachverhalts verwendet wird, in der Konzeption aber von den gleichen Prämissen ausgeht.

Das Beispiel mit der Münze können wir zur Konstruktion des Ereignisbaumes verwenden:

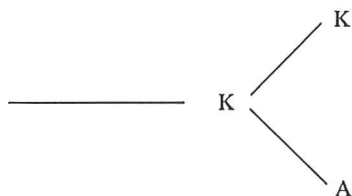


Aus dem Ereignisbaum können wir nun die einzelnen möglichen Ausgänge, die den Ereignisraum ausmachen, ablesen. Wenn wir entlang der Äste von innen nach außen gehen, erhalten wir die einzelnen möglichen Ereignisse, z.B. (K, K, K), (K, K, A), (K, A, K) usw. An dieser Stelle können wir auch darauf hinweisen, daß ein solcher Baum in seiner speziellen Ausprägung als binärer Baum (Huffman-Baum) in der

Informationstheorie zur Veranschaulichung der binären Codierung benutzt wird. Es wird an anderer Stelle ausführlich darüber zu sprechen sein.

Mit dem Ereignisbaum können wir auch den Fall der bedingten Ereignisse gut verdeutlichen.

Wenn wir eine Verzweigung mit den dazugehörigen Ästen ausschneiden, die davor liegenden Ereignisse außer Acht lassen und diesen Ausschnitt als einen neuen Baum betrachten, dann sind die Ereignisse der Äste des neuen Baumes bedingte Ereignisse, z.B. ein einfaches Beispiel aus dem Ereignisbaum des Münzenwurfs.



Wir haben hier zwei bedingte Ereignisse

$$K|K \quad \text{und} \quad A|K$$

d.h. K gegeben K und A gegeben K .

Zwei Ereignisse A_1 und A_2 sind unabhängig, wenn die Wahrscheinlichkeit des Eintretens des Ereignisses A_1 nicht davon beeinflusst wird, ob das Ereignis A_2 schon eingetreten ist oder nicht. D.h. aber, seine bedingte ist gleich mit seiner unbedingten Wahrscheinlichkeit.

$$P(A_1) = P(A_1 | A_2)$$

In diesem Fall werden wir sagen:

A_1 ist *stochastisch* unabhängig von A_2 .

Diese Definition gilt auch wenn $P(A_1) = 0$; in diesem Fall ist

$P(A_1 | A_2)$ nicht definiert.

Die Bezeichnung stochastisch unabhängig ist synonym zu der Bezeich-

nung statistisch unabhängig.

Wenn wir jetzt $P(A_1)$ in die Formel des Multiplikationssatzes einsetzen, erhalten wir

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

Die Wahrscheinlichkeit des gleichzeitigen Eintretens ist infolgedessen bei unabhängigen Ereignissen (A_1 und A_2) gleich dem Produkt ihrer Einzelwahrscheinlichkeiten.

Mit Hilfe der bedingten Wahrscheinlichkeit können wir nun den Multiplikationssatz auf mehrere Ereignisse erweitern, und erhalten

$$P(A_1 \cap A_2 \cap \dots \cap A_n) =$$

$$P(A_1) P(A_2 | A_1) P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

Die Erweiterung der Konzeption der Unabhängigkeit auf mehrere Ereignisse stößt auf gewisse Schwierigkeiten, auf die wir hier nicht ausführlich einzugehen brauchen. Mehrere Ereignisse sind *vollständig* unabhängig, wenn folgende zwei Beziehungsarten bestehen:

a) $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$, $P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$ usw.

d.h. wenn die Ereignisse paarweise unabhängig sind. Diese Bedingung genügt aber nicht; es muß noch gelten

b) $P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$.

Diese Bedingung folgt nicht aus der Bedingung a) d.h. aus der paarweise Unabhängigkeit von Ereignissen folgt nicht, daß zwischen den Ereignissen keine Abhängigkeit besteht. Bedingung b) ist infolgedessen die Definition der vollständig unabhängigen Ereignisse.

Satz von der totalen Wahrscheinlichkeit. Wenn die zufälligen Ereignisse A_1, A_2, \dots, A_n paarweise disjunkt sind und die Menge der Elementarereignisse ausschöpfen, dann gilt für ein zufälliges Ereignis B

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B)$$

wobei sich die Ereignisse $A_i \cap B \mid i = 1 \dots n$ gegenseitig ausschließen.

Wenn wir jetzt den Multiplikationssatz anwenden, erhalten wir

$$\begin{aligned} P(B) &= P(A_1) \cdot P(B | A_1) + P(A_2) \cdot P(B | A_2) + \dots \\ &\quad + P(A_n) \cdot P(B | A_n) \\ &= \sum_{i=1}^n P(A_i) \cdot P(B | A_i) \end{aligned}$$

(wenn für alle i : $P(A_i) \neq 0$).

Unter diesen Voraussetzungen läßt sich nun auch das Bayessche Theorem, das eine Beziehung zwischen verschiedenen bedingten Wahrscheinlichkeiten enthält, formulieren.

Wenn $\{A_1, A_2, \dots, A_n\}$ eine Menge von n sich gegenseitig ausschließenden Ereignissen ist und B ein Ereignis mit $B \subseteq A_1 \cup A_2 \cup \dots \cup A_n$

damit ist auch $P(B) < \sum_{i=1}^n P(A_i)$, dann gilt

$$\begin{aligned} P(A_i | B) &= \frac{P(B | A_i) \cdot P(A_i)}{P(B | A_1) \cdot P(A_1) + P(B | A_2) \cdot P(A_2) + \dots + P(B | A_n) \cdot P(A_n)} \\ &= \frac{P(B | A_i) \cdot P(A_i)}{\sum_{i=1}^n P(B | A_i) \cdot P(A_i)} \end{aligned}$$

Das Bayessche Theorem dient zur Bestimmung der a posteriori Wahrscheinlichkeit. Dabei wird $P(A_i | B)$ als die a posteriori Wahrscheinlichkeit für A_i , nachdem B eingetreten ist, interpretiert; $P(A_i)$ ist die a priori Wahrscheinlichkeit für A_i .

“Das Bayessche Theorem transformiert also die A-priori-Wahrscheinlichkeit eines Ereignisses im Lichte der Beobachtung B in eine A-posteriori-Wahrscheinlichkeit”⁴³⁾

Unabhängige Versuche. R sei ein endlicher Wahrscheinlichkeitsraum mit den Punkten (Elementen) $P_1, P_2 \dots P_n$. Unter s unabhängigen Versuchen bezüglich R verstehen wir den Wahrscheinlichkeitsraum T , der aus geordneten s -tupeln von Elementen von R besteht, wobei die Wahrscheinlichkeit eines s -tupels durch das Produkt der Wahrscheinlichkeiten seiner Komponenten gegeben wird:

$$P(A_{i_1}, A_{i_2}, \dots, A_{i_s}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_s}) \quad .$$

Ein solcher Prozess wiederholter Versuche wird ein stochastischer Prozess genannt.

Die Definition unabhängiger Versuche, die wir oben angeführt haben, impliziert folgenden Tatbestand:

Bei einem stochastischen Experiment sind unter Zugrundelegung eines bestimmten Merkmals n verschiedene Ereignisse (als Ausgänge des Experiments) möglich. Dieses Experiment wird s -mal durchgeführt, und die s realisierten Ausgänge bilden in ihrer zeitlichen Realisationsfolge ein Elementarereignis $(A_{i_1}, A_{i_2}, \dots, A_{i_s})$ des stochastischen Prozesses (d.h. des s -fachen Experiments).

Das s -tupel $(A_{i_1}, A_{i_2}, \dots, A_{i_s})$ bezeichnen wir eine *Stichprobe*. Diese Konzeption entspricht einer analogen Anordnung, bei der die Werte A an s Merkmalsträgern beobachtet werden, die aus einer Grundgesamtheit entnommen werden.

Zur Bestimmung der Wahrscheinlichkeit einer solchen Stichprobe ist die Errechnung der Gesamtzahl aller möglichen Stichproben notwendig. Solche Abzählverfahren durch bestimmte Definitionen festgesetzter Anordnungen von Elementen einer endlichen Menge werden durch die Kombinatorik geliefert.

1. Kombinatorik

Gegenstand der Kombinatorik ist die Berechnung der Anzahl möglicher Anordnungen von n verschiedenen Dingen oder einer Auswahl davon.

Diese Dinge, allgemein Elemente genannt, sind Teile einer statistischen

Gesamtheit, die sachlich, örtlich und zeitlich definiert sein muß. Was diese Elemente im einzelnen sind, hängt vom Untersuchungsgegenstand ab. Es kann sich um Wörter, Buchstaben, Zahlen, Bücher, Personen, Faktoren usw. handeln.

Die Grundlegung der Beweisführung in der Kombinatorik erfolgt durch zwei Regeln, die entweder einzeln oder zusammen zur Anwendung kommen:

Summenregel: wenn ein Element A auf m verschiedene Wege zur Auswahl kommen kann und ein Element B auf n andere verschiedene Wege, dann können "entweder A oder B" auf $m + n$ verschiedene Wege zur Auswahl kommen.

Produktregel: Wenn ein Element A auf m verschiedene Wege zur Auswahl kommen kann und danach B auf n verschiedene Wege, dann können "A und B" bei dieser Anordnung (Abfolge) auf $m \cdot n$ Wege zur Auswahl kommen.

In der ersten Regel ist die Bedingung impliziert, daß sich die Auswahlmöglichkeiten von A und B gegenseitig ausschließen, d.h. daß es nicht möglich ist, daß A und B auf dem gleichen Wege ausgewählt werden. Die zweite Regel erhält meistens dann ihre Gültigkeit, wenn die Auswahlen unabhängig voneinander sind, d.h. wenn die Anordnung der Elemente ohne Belang bleibt.

Grundsätzlich können wir 4 Kriterien zur speziellen Charakterisierung der verschiedenen Anordnungen unterscheiden, wobei wir die Buchstaben als Repräsentationssymbole der gegebenen Elemente auffassen:

- 1) Sind die einzelnen Elemente *verschieden* (A, B, C, D, E, . . .) oder sind einige Elemente *gleich* (A, B, B, C, D, E, E, . . .) ?
- 2) Soll die *Reihenfolge* berücksichtigt werden (B, C, \neq C, B) oder nicht (B, C = C, B) ?
- 3) Sollen einige der Elemente *ausgewählt* werden (k aus n) oder nicht?
- 4) Soll *Wiederholung* (Zurücklegen der gewählten Elemente in die Gesamtheit) möglich sein (A, A) oder nicht (A, A) ?

Unter Zugrundelegung dieser Kriterien kommen wir zu folgenden Definitionen:

- 1.a. Jede Anordnung einer endlichen Zahl von Elementen, die dadurch entsteht, daß man sämtliche Elemente in irgendeiner Reihenfolge nebeneinandersetzt, nennen wir eine *Permutation* dieser Elemente. (Berücksichtigung der Reihenfolge).
Die Gesamtzahl der Permutationen von n Elementen beträgt

$$P_{(n)} = n! \quad (a)$$

- b. Sind unter den n Elementen n_1, n_2, \dots, n_k die Anzahl gleicher Elemente, so beträgt die Gesamtzahl der Permutationen

$$P_{n_1, n_2, \dots, n_k}(n) = \frac{n!}{n_1! n_2! \dots n_k!} \quad (b)$$

Wir wollen hier paradigmatisch die Ableitung der Formel für die Gesamtzahl der Permutationen vorführen. Beim jeweiligen n haben wir die entsprechenden Permutationen angeführt. (Bei $n = 1$ ist $P_{(1)} \doteq 1$)

$n = 2$	$n = 3$	$n = 4$			
AB	ABC	ABCD	BACD	CABD	DABC
BA	ACB	ABDC	BADC	CADB	DACB
	BAC	ACBD	BCAD	CBAD	DBAC
	BCA	ACDB	BCDA	CBDA	DBCA
	CAB	ADBC	BDAC	CDAB	DCAB
	CBA	ADCB	BDCA	CDBA	DCBA
<hr/>		<hr/>			
$P_{(2)} = 2$	$P_{(3)} = 6$	$P_{(4)} = 24$			

Aus dieser *lexikographischen Anordnung* erkennen wir bestimmte Gesetzmäßigkeiten: Das erste Element (es ist völlig gleichgültig, welches der Elemente wir als erstes bezeichnen) können wir so oft an die erste Stelle einer Permutation setzen, wie die übrigen $n-1$ Elemente permu-

tieren lassen, d.h. $P_{(n-1)}$ -mal. So oft können wir wiederum das zweite Element an die erste Stelle setzen, d.h. wieder $P_{(n-1)}$ -mal.

Allgemein kann man dann sagen: n Elemente lassen sich immer genau n -mal so oft permutieren wie $n-1$ Elemente. Daraus ergibt sich die Rekursionsformel

$$P_{(n)} = n \cdot P_{(n-1)}$$

Aus dieser Formel ergibt sich

$$P_{(2)} = 2 \cdot P_{(1)} = 2 \cdot 1 = 2 !$$

$$P_{(3)} = 3 \cdot P_{(2)} = 3 \cdot 2 \cdot 1 = 3 !$$

$$P_{(4)} = 4 \cdot P_{(3)} = 4 \cdot 3 \cdot 2 \cdot 1 = 4 !$$

.....

$$P_{(n)} = n P_{(n-1)} = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1 = n !$$

Beispiel für 1b):

Für $n=3$ Elemente A, A, B sind folgende verschiedene Fälle zu unterscheiden

AAB nach der Formel (1 b) haben wir

ABA

BAA

$$P_2(3) = \frac{3 !}{2 !} = 3$$

- 2.a. Jede Auswahl von k Elementen aus allen gegebenen n Elementen, wobei es auf die Reihenfolge der Elemente nicht ankommt, (A, B = B, A) nennen wir eine *Kombination* dieser Elemente. Die Gesamtzahl der Kombinationen der Klasse k von n Elementen (auch: von n Elementen zur k -ten Klasse) ohne Wiederholung beträgt:

$$C_k(n) = \binom{n}{k} = \frac{n!}{k! (n-k)!} \quad (a)$$

Beispiel: Für $n = 3$ Elemente A, B, C, von denen $k = 2$ ausgewählt werden, sind folgende Fälle zu unterscheiden:

A, B

$$A, C \quad C_2(3) = \frac{3}{2} = \frac{3!}{2! (3-2)!} = 3$$

B, C

- b. Unter einer Kombination der Klasse k von n Elementen *mit Wiederholung* verstehen wir jede Auswahl von insgesamt k Elementen aus allen gegebenen Elementen, bei der jedes Element so oft vorkommen kann, wie es die Klassenzahl zulässt. Die Gesamtzahl der Kombinationen der Klasse k von n Elementen mit Wiederholung beträgt:

$$C'_k(n) = \binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)! k!} \quad (b)$$

Beispiel: Für $n = 3$ Elemente A, B, C, von denen $k = 2$ ausgewählt werden, sind folgende Fälle zu unterscheiden:

A, B,

A, C,

B, C

A, A

B, B

C, C

$$C'_2(3) = \binom{3+2-1}{2} = \frac{4!}{2! 2!} = 6$$

- 3.a. Jede Auswahl von k Elementen aus allen gegebenen n Elementen, wobei zwei Auswahlen, die sich nur durch die Anordnung ihrer Elemente unterschieden, als verschieden anzusehen sind (A, B \neq B, A), nennen wir eine *Variation* dieser Elemente.

Die Gesamtzahl aller Variationen der k-ten Klasse von n Elementen ohne Wiederholung beträgt:

$$V_k(n) = \frac{n!}{(n-k)!} \quad (a)$$

Beispiel: Für $n = 3$ Elemente A, B, C, von denen $k = 2$ ausgewählt werden, sind folgende verschiedene Fälle zu unterscheiden:

A, B

B, A

A, C

C, A

B, C

C, B

$$V_2(3) = \frac{3!}{(3-2)!} = 6$$

- b. Die Gesamtzahl aller Variationen der k-ten Klasse von n Elementen mit Wiederholung beträgt:

$$V'_k(n) = n^k \quad (b)$$

Beispiel: Für $n = 3$ Elemente A, B, C, von denen $k = 2$ ausgewählt werden, sind folgende verschiedene Fälle zu unterscheiden:

A, A

B, B

C, C

A, B

B, A

A, C

C, A

B, C

C, B

$$V'_2(3) = 3^2 = 9$$

Die Variationen nennt man auch Permutationen, so daß wir im Falle 3.a. von Permutationen von k Elementen aus n verschiedenen Elementen ohne Wiederholung, und im Falle 3.b. von Permutationen von k Elementen aus n Elementen mit Wiederholung sprechen könnten. Wir behalten aber die hier eingeführte Dreiteilung bei, da sie u.E. übersichtlicher und weniger verwechslungsanfällig ist.

Wenn wir diese Zweiteilung zugrundelegen, können wir die Definitionen folgendermaßen formulieren:

1. Definition. Eine Permutation von n Elementen zur k -ten Klasse ist eine geordnete Anordnung oder Auswahl von k dieser n Elementen
2. Definition. Eine Kombination von n Elementen zur k -ten Klasse ist eine Anordnung oder Auswahl von k dieser n Elemente ohne Berücksichtigung der Ordnung.

Sind von n Elementen k und $(n - k)$ jeweils gleich, so gibt:

$$P_{k, (n-k)}(n) = C_k(n)$$

Aus diesen allgemeinen Definitionen lassen sich unter Berücksichtigung zusätzlicher Kriterien weitere Lehrsätze ableiten. Ohne Anspruch auf Vollständigkeit und zwecks beispielhafter Demonstration möglicher Erweiterungen seien hier einige Lehrsätze angeführt:

- 2.1. Die Anzahl T derjenigen unter den Kombinationen der k -ten Klasse von n Elementen, die q vorgeschriebene Elemente ($1 \leq q \leq k$) enthalten, ist der Anzahl der Kombination der $(k - q)$ -ten Klasse von $(n - q)$ Elementen gleich.

$$T = C_{k-q}(n-q) = \binom{n-q}{k-q}$$

$$\text{Mit Wiederholung } T' = C'_{k-q}(n) = \binom{n+k-q-1}{k-q}$$

- 2.2. Die Anzahl S derjenigen unter den Kombinationen der k -ten Klasse von n Elemente, die keines von q vorgeschriebenen Elementen ($1 \leq q \leq n$) enthalten, ist der Anzahl der Kombina-

tionen der k-ten Klasse von $(n-q)$ Elementen gleich.

$$S = C_k (n-q) = \binom{n-q}{k}$$

Mit Wiederholung

$$S' = C'_k (n-q) = \binom{n+k-q-1}{k}$$

- 2.3. Die Anzahl R derjenigen unter den Kombinationen der k-ten Klasse von n Elementen, die mindestens eines von q vorgeschriebenen Elementen ($1 \leq q \leq n$) enthalten, ist

$$R = C_k (n) - C_k (n-q) = \binom{n}{k} - \binom{n-q}{k}$$

Mit Wiederholung

$$R' = C'_k (n) - C'_k (n-q) = \binom{n+k-1}{k} - \binom{n+k-q-1}{k}$$

- 3.1. Die Anzahl n derjenigen unter den Variationen ohne Wiederholung der k-ten Klasse von n Elementen, die q vorgeschriebene Elemente ($0 \leq q \leq k$) enthalten, ist

$$M = k! \cdot C_{k-q} (n-q)$$

- 3.2. Die Anzahl N derjenigen unter den Variationen ohne Wiederholung der k-ten Klasse von n Elementen, die keines von q vorgeschriebenen Elementen ($0 \leq q \leq n$) enthalten, ist

$$N = V_k (n-q) = \frac{(n-q)!}{(n-q-k)!}$$

Mit Wiederholung

$$N' = V'_k (n-q) = (n-q)^k.$$

- 3.3. Die Anzahl Q derjenigen unter den Variationen ohne Wiederholung der k -ten Klasse von n Elementen, die mindestens eines von q vorgeschriebenen Elementen ($0 \leq q \leq n$) enthalten, ist

$$Q = V_k(n) - V_k(n-q) = \frac{n!}{(n-k)!} - \frac{(n-q)!}{(n-q-k)!}$$

Mit Wiederholung

$$Q' = V'_k(n) - V'_k(n-q) = n^k - (n-q)^k$$

8. Nicht-klassische Statistik.

Wir haben im ersten Teil den Begriff der Verteilung im Zusammenhang mit beobachteten Ereignissen kennengelernt: sie entsprach dem allgemeinen Sprachgebrauch und bedeutete die Zuteilung irgendwelcher Objekte (hier: Beobachtungswerte, Versuchsausgänge) zu irgendwelchen Klassen oder Kategorien (hier: Ereignisse).⁴⁴⁾

Wir wollen nun hier eine Konkretisierung und Differenzierung vornehmen, die zum Verständnis der weiteren Ausführungen in diesem Abschnitt beitragen wird. Eine Verteilung wird hierbei als eine Zuordnung von Objekten an Kästchen oder Zellen beschrieben. Die Objekte können in ihrer Art und Anzahl beliebig sein und die Zellen können in ihrer Art, Anzahl und Kapazität frei spezifiziert werden, wobei die Anordnung von Objekten innerhalb einer Zelle (im Sinne der Kombinatorik) berücksichtigt werden oder belanglos bleiben kann. Wenn die Anzahl der Zuordnungen im Mittelpunkt der Diskussion steht, sprechen wir von einem Verteilungsproblem, wenn dagegen die Anzahl der Objekte innerhalb gegebener oder willkürlicher Zellen von Interesse ist, haben wir mit einem Problem der Besetzung zu tun. Die Ähnlichkeit mit der Fragestellung der Kombinatorik ist unverkennbar, der Gesichtspunkt ist aber verschieden.⁴⁵⁾

Wir haben bis jetzt die Ermittlung der Wahrscheinlichkeit auf die grundsätzliche Konzeption von gleichwahrscheinlichen Ereignissen gestützt. Diese Vorstellung der Gleichwahrscheinlichkeit, die die Grundlage der klassischen Wahrscheinlichkeitstheorie bildet, ist aber, unabhängig von der logischen Zirkularität der Wahrscheinlichkeitsdefinition, die durch die Voraussetzung der Gleichwahrscheinlichkeit gewisser Ereignisse bedingt ist, nicht immer eindeutig. Das Bترandsche Paradoxon (das wir hier nicht zu behandeln brauchen) zeigt, daß diese Vorstellung variierbar ist, und daß man je nach der konkreten Vorstellung zu unterschiedlichen Ergebnissen gelangt.

Die drei Wahrscheinlichkeitsmodelle, die hier behandelt werden, sind aus Diskussionen in der Physik entstanden, konkret: aus der Unzulänglichkeit des klassischen Wahrscheinlichkeitsmodells ge-

wisse Vorgänge im kernphysikalischen Bereich adäquat zu beschreiben. Man geht davon aus, daß n Zellen (Kästchen) mit r Objekten (Teilchen, Partikeln, Bällen) besetzt werden sollen.

Je nach den Voraussetzungen, die den experimentellen Rahmen dieses Vorhabens bestimmen, ergeben sich drei verschiedene Modelle.

8.1 Das Boltzmann-Maxwell-Modell.

Die Voraussetzung dieses klassischen Modells ist, daß die einzelnen zu verteilenden Objekte individuell identifizierbar, d.h. unterscheidbar sind. Jedes Objekt hat n Auswahlmöglichkeiten. Es gibt n^r Möglichkeiten die r Objekte auf n Zellen zu verteilen, die gleichwahrscheinlich sind. (Es entspricht den Variationen mit Wiederholung). Wenn wir die Anzahl der Objekte in der i -ten Zelle r_i bezeichnen, dann beschreibt das n -tupel

$$r_1 + r_2 + \dots + r_n = r$$

eine mögliche Konfiguration der Besetzungszahlen. Die Wahrscheinlichkeit die gegebenen Besetzungszahlen r_i zu erreichen ist dann

$$P_{B-M} = \frac{1}{n^r} \cdot \frac{r!}{r_1! \cdot r_2! \cdot \dots \cdot r_n!}$$

Mit $r = 2$ und $n = 3$ ergibt sich folgendes Bild der Besetzungsmöglichkeiten.

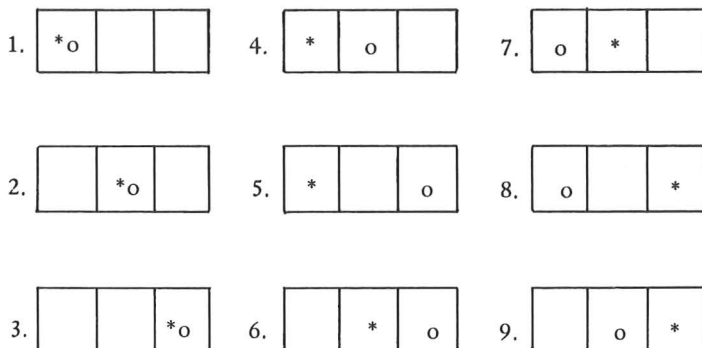


Fig. 10

Nun müssen wir eine spezifische Besetzungskonfiguration zugrundelegen, z.B. $r_1 = 1$ und $r_3 = 1$, $r_1 + r_3 = 2$ ($r_2 = 0$)
 Nach der Formel ergibt sich als Wahrscheinlichkeit einer solchen Konfiguration

$$P_{B-M} = \frac{1}{3^2} \cdot \frac{2!}{1! 1!} = \frac{2}{9}$$

Wie man aus der Fig. sieht erfüllen bei 9 gleichwahrscheinlichen Möglichkeiten zwei die obige Besetzungsspezifikation, sodaß die Wahrscheinlichkeit bei zwei "günstigen" Ereignissen und neun möglichen tatsächlich $2/9$ beträgt. Die "günstigen" Ereignisse sind die Fälle 5 und 8.

Wie man hier erkennt, liegt der Vorstellung der Gleichwahrscheinlichkeit in diesem Modell eine ganz bestimmte Konzeption zugrunde. Wiederholte Versuche sind unternommen worden, die darauf abzielten, den Beweis zu erbringen, daß das Verhalten der physikalischen Elementarteilchen dem Boltzmann-Maxwell-Modell entspricht, d.h. durch dieses Modell beschrieben werden kann. Die moderne Physik hat aber gezeigt, daß dieses Modell auf keine der bekannten Elementarteilchen angewendet werden kann, denn in keinem Fall konnte man auch nur eine annähernde Gleichwahrscheinlichkeit aller r^n Konfigurationen feststellen. Es wurden daher zwei Wahrscheinlichkeitsmodelle eingeführt, die jeweils auf einen Teil des durch das klassische Modell nicht beschreibbaren atomaren Bereichs angewendet werden können.

8.2 Das Bose-Einstein-Modell

Dieses Modell geht davon aus, daß die Objekte, mit denen n Zellen zu besetzen sind, individuell nicht identifizierbar sind. Man sieht also von den individuellen Objekten völlig ab und betrachtet nur Besetzungszustände, die allein unterscheidbar und gleichwahrscheinlich sind. Zwei Verteilungen sind in diesem Fall unterscheidbar, nur wenn die entsprechenden Konfigurationen der Besetzungszahlen, d.h. die n -tupeln (r_1, r_2, \dots, r_n) nicht identisch sind. (Bei identischen n -tupeln sind die Verteilungen gleich, auch wenn in einzelnen Zellen die

einzelnen Objekte nicht identisch sind).

Die Anzahl der unterscheidbaren Verteilungen ergibt sich aus der Formel für Kombinationen mit Wiederholung. (Aus n Zellen werden Zellen ausgewählt, ihren Besetzungszahlen entsprechend), d.h.

$$\binom{n+r-1}{r} = \binom{n+r-1}{n-1}$$

Die Wahrscheinlichkeit, die jeder dieser Verteilung zugeordnet wird, ist daher

$$P_{B-E} = \frac{1}{\binom{n+r-1}{r}} = \frac{r! (n-1)!}{(n+r-1)!}$$

Mit $r = 2$ und $n = 3$ ergeben sich folgende unterscheidbare Verteilungen.

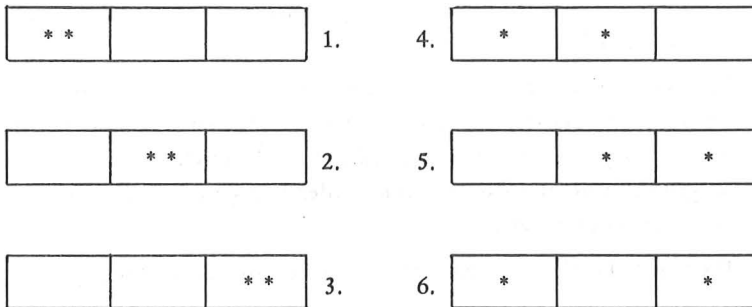


Fig. 11

Wie man sieht sind die Fälle 4 und 7, 5 und 8 und 6 und 9 des klassischen Modells nun mehr nicht mehr unterscheidbar und schmelzen hier jeweils zu einem Fall zusammen.

Alle sechs Verteilungen sind gleichwahrscheinlich und die Wahrscheinlichkeit der Realisierung einer solchen Verteilung ist gleich $1/6$. Wir

können es auch aus der Formel errechnen und erhalten

$$P_{B-E} = \frac{2! (3-1)!}{(3+2-1)!} = \frac{1}{6}$$

Die Änderung der Konzeption, die der Vorstellung der Gleichwahrscheinlichkeit zugrundeliegt, ändert auch das ganze Modell und seine Ergebnisse. Das Bose-Einstein-Modell kann zum Studium von Photonen, Atomkernen und Atomen, die eine gerade Anzahl von Elementarteilchen enthalten angewendet werden.⁴⁶⁾

Die r nicht unterscheidbaren Objekte sind dann die in Betracht gezogenen Teilchen und die n verschiedenen Zellen verschiedene mögliche und unterscheidbare Energiezustände.

8.3 Das Fermi-Dirac-Modell.

Diesem Wahrscheinlichkeitsmodell liegen folgende zwei Hypothesen zugrunde: (1) es ist unmöglich, daß mehr als ein Objekt, Teilchen, die gleiche Zelle besetzen (In der Atomphysik bekannt als "Pauli-Verbot"). Daraus folgt, daß $r \leq n$ sein muß; (2) alle Verteilungen, die die erste Hypothese befriedigen sind gleichwahrscheinlich.

Eine Verteilung wird dann durch die Feststellung vollständig beschrieben, welche der n Zellen ein Objekt enthalten, und weil die Anzahl der Objekte r ist, können die Zellen in $\binom{n}{r}$ verschiedenen Weisen ausgewählt werden (Aus n Zellen werden r ausgewählt - und mit einem Objekt besetzt).

Die Wahrscheinlichkeit des Eintretens einer solchen Verteilung ist

$$P_{F-D} = \frac{1}{\binom{n}{r}} = \frac{r! (n-r)!}{n!}$$

Mit $r = 2$ und $n = 3$ haben wir in diesem Modell folgende unterscheidbare Verteilungen.



Fig. 12

Wir haben hier drei gleichwahrscheinliche unterscheidbare Verteilungen, und die Wahrscheinlichkeit der Realisierung einer dieser Verteilungen ist dann $1/3$, oder wenn wir in die Formel einsetzen

$$P_{F-D} = \frac{2! (3-2)!}{3!} = \frac{1}{3}$$

Das Fermi-Dirac-Modell kann zum Studium von Protonen, Neutronen und Elektronen (bzw. Atomen mit einer ungeraden Anzahl von Elementarteilchen) erfolgreich angewendet werden.

8.4 Zusammenfassende Interpretation.

Aus diesen Ausführungen über drei bekannte Wahrscheinlichkeitsmodelle läßt sich nunmehr sowohl die Problematik einer impliziten und unreflektierten Zugrundelegung gleichwahrscheinlicher Ereignisräume als auch die Flexibilität dieser Konzeption im Hinblick auf ihre Anwendungsmöglichkeiten in einer breiten Skala von wissenschaftlichen Fragestellungen erkennen. Die Berechtigung des einen oder des anderen Modells erwächst aus dem jeweiligen Erfolg, eine wissenschaftlich formulierte Situation adäquat zu beschreiben; keiner dieser Ansätze kann für sich Allgemeingültigkeit beanspruchen. Die Wahl zu Gunsten eines dieser Modelle und ihre Rechtfertigung kann nicht durch eine a priori Argumentation begründet werden. Es gibt z.B. im Falle der Physik kein Postulat der reinen Vernunft, das verbietet Protonen und Photonen nach den gleichen Wahrscheinlichkeitsgesetzmäßigkeiten zu behandeln. ⁴⁷⁾

Eine grobe aber instruktive Interpretation der Unterschiede zwischen diesen drei Wahrscheinlichkeitsmodellen hat der englische Mathematiker E. Whittaker gegeben. ⁴⁸⁾

Ein Zug steht an einem Bahnsteig und Passagiere steigen ein. Wenn

wir davon ausgehen, daß jeder Passagier sein Abteil ganz zufällig (blindlings) aussucht, d.h. ohne davon Notiz zu nehmen, ob andere Passagiere in dieses Abteil eintreten oder nicht, dann ist es möglich, die Wahrscheinlichkeit jeder speziellen Verteilung der Passagiere auf die Abteile zu ermitteln. Diese Anordnung würde dem Boltzmann-Maxwell-Modell entsprechen.

Wenn wir andererseits von der Annahme ausgehen, daß die Passagiere ein Abteil für sich allein bevorzugen, oder daß sie auf jeden Fall Fenstersitze haben wollen, dann würden sich die Wahrscheinlichkeiten für die verschiedenen Verteilungen der Passagiere auf die Abteile ändern. Die Wahrscheinlichkeit würde für solche Verteilungen steigen, für welche jedes Abteil mindestens von einem Passagier besetzt und kein Abteil überfüllt ist. Diese Anordnung würde dem Fermi-Dirac-Modell am nächsten sein.

Wenn wir jetzt die entgegengesetzte Annahme machen, daß die Passagiere sehr gesellige Menschen sind, die nicht allein reisen wollen und deswegen Abteile aufsuchen, in denen auch andere Menschen hineingehen, dann würden sich die Wahrscheinlichkeiten für die verschiedenen Verteilungen wieder ändern und diesmal zugunsten solcher Verteilungen, bei denen einige Abteile voll besetzt und andere leer wären. Diesen Fall würde man dem Bose-Einstein-Modell zurechnen.

Das Fermi-Dirac- und das Bose-Einstein-Modell weichen vom klassischen Modell in entgegengesetzten Richtungen ab.⁴⁹⁾

9. Informationstheorie

9.1 Der informationstheoretische Ansatz

Mit dem Erscheinen der zwei grundlegenden Arbeiten von C. E. Shannon und N. Wiener im Jahre 1948 setzte eine Entwicklung ein, die den informationstheoretischen Ansatz zum Untersuchungs- und Diskussionsgegenstand mehrerer wissenschaftlicher Disziplinen machte.

Das linguistische Interesse für die Informationstheorie gründet in der Tatsache, daß die natürliche Sprache eines der wichtigsten Mittel zur

Übertragung von Informationen ist.

Die Informationstheorie betrachtet die natürliche Sprache als ein komplexes System der Nachrichtenkodierung, wobei verschiedene Eigenschaften dieses Systems durch die Heranziehung statistischer Methoden studiert werden.

Die Entwicklung der Theorie hat sich in einem Aufschwung der Forschungstätigkeit auf verschiedenen Gebieten, in einer Erweiterung der ursprünglichen Konzeptionen und in einer Ausweitung des Objektbereichs über die Grenzen der Statistik hinaus niedergeschlagen. Hier wird allerdings eine Einschränkung auf die statistische Theorie der selektiven Information und auf die wichtigsten, damit zusammenhängenden Anwendungen auf die natürlichen Sprachen notwendig sein. Die Informationstheorie beschäftigt sich mit dem Prozeß der Informations- (Nachrichten-) Übertragung mittels eines Kommunikationssystems, wobei dies eine abstrakte Grundstruktur ist, die vielfältige konkrete Realisationsformen erhalten kann. Damit ist jedes natürliche oder konventionelle Zeichensystem, das die Übertragung von Informationen von einem kybernetischen System organischer, technischer, gesellschaftlicher Art zum anderen übernimmt aus informationstheoretischer Sicht eine Sprache.⁵⁰⁾

Der Begriff der Information im Sinne des natürlich- sprachlichen Gebrauchs muß aber hier aufgegeben werden; die zwischen diesem letzten Begriff und dem in der Kommunikationsforschung verwendeten bestehen zwar gewisse Beziehungen. Es ist aber einfacher davon abzusehen und der neuen theoretischen Begriffsbildung zu folgen, bzw. sie tentativ anzugehen.

Der statistischen Theorie der Information liegt ein Modell zugrunde, die abstrakte Kommunikationsstruktur, dessen charakteristische Eigenschaften den Objektbereich umschreibt und abgrenzt, wobei die frühere technische Bezogenheit dieses Modells zu Gunsten einer viel allgemeineren Interpretation aufgehoben wurde. Die abstrakte Kommunikationsstruktur, die dem informationstheoretischen Ansatz zugrunde liegt, besteht aus fünf verschiedenen Elementen.

- (1) Informationsquelle (Nachrichtenquelle). Diese ist der Teil des Systems, der die zu übertragende Nachricht erzeugt.
- (2) Sendewandler. Dieser transformiert die Nachricht in übertragbare Signale und leitet sie dem Übertragungsteil zu.
- (3) Übertragungsteil. Dieser übernimmt die Überführung der transformierten Nachricht über die räumliche Entfernung zum Empfänger.
- (4) Empfangswandler. Dieser wandelt die empfangenen Signale in primäre Nachricht um, und leitet sie dem
- (5) Empfänger zu, für welchen die übertragene Nachricht bestimmt ist.



Die Funktion des Kommunikationssystems besteht darin, eine einmal ausgewählte Nachricht einem bestimmten Empfänger zuzuführen. Nicht jeder Aspekt einer Nachricht wird in dem informationstheoretischen Ansatz berücksichtigt. Die inhaltliche Substanz einer Information, ihre Wichtigkeit u. ä. m. bleibt in diesem Zusammenhang außer Acht; nur eine quantitative Beschreibung der Information, die aus einer Nachrichtenquelle ausgeht, steht hier im Mittelpunkt. Durch die Ausschaltung der semantischen Aspekte der Kommunikation ist die entscheidende Möglichkeit geschaffen worden, quantitative Vergleiche zwischen verschiedenen Nachrichten anzustellen, und zwar Vergleiche hinsichtlich einer von diesem Standpunkt aus gesehen wichtigen Eigenschaft, nämlich ihrer statistischen Struktur. ⁵¹⁾ Dadurch sind die Nachrichten, die hier behandelt werden, irgendwelche Sequenzen von unterscheidbaren Zeichen, wie z.B. telegraphierte Nachrichten, sprachliche Nachrichten usw. . Die Anzahl der unterscheidbaren Zeichen, die zur Bildung einer Nachricht herangezogen werden können, d.h. das Zeichenrepertoire einer Quelle, ist eine endliche Menge, die als Zeichenvorrat

oder Alphabet bezeichnet wird.

Die wichtigste Eigenschaft einer Information ist darin zu sehen, daß sie etwas Neues vermittelt, d.h. etwas was nicht mit Bestimmtheit vorhergesagt werden kann; denn wenn sie vollständig vorhersagbar wäre, dann bestünde für den Empfänger keine Veranlassung diese Nachricht aufzunehmen, um zu wissen, was sie aussagt. "Die Informationstheorie der Nachrichtentechnik führt der Begriff der 'Information', so mathematisch sie ihn auch präzisieren mag, durchaus anschaulich ein. Gemäß ihren Vorstellungen ist 'Information' stets 'Neuigkeit', 'Innovation', also das originale Moment in einem Ordnungsschema. Je reiner in dieser Hinsicht die 'Information' ist, desto unvorhersehbarer, desto unwahrscheinlicher ist sie, einen desto geringeren Häufigkeitscharakter hat sie dann. Doch handelt es sich nie um den bedeutungshaften, inhaltlichen Sinn der 'Information', sondern lediglich um den numerischen Betrag der 'Innovation'.⁵²⁾

Information besteht also in der ungewissen Auswahl einer Nachricht aus einer Menge möglicher Nachrichten, d.h. bevor eine Informationsquelle ein Zeichen weiterleitet herrscht Ungewißheit darüber, welches spezielle Zeichen aus der Menge aller möglichen Zeichen ausgewählt wird. Aus dieser einfachen Konzeption ergibt sich, daß die Information, die ein Zeichen trägt um so größer ist, je größer die Ungewißheit seines Auftretens ist, oder anders ausgedrückt: mit steigender Eintrittswahrscheinlichkeit eines Zeichens nimmt sein Informationsgehalt ab.

Indem man aber dieses informationstheoretische Selektionsprinzip aus einer Menge von Alternativen (Zeichenvorrat) zugrundelegt, kommt man zu der Erkenntnis, daß die Probleme der Informationstheorie statistischer Natur sind.⁵³⁾

Zur Bildung einer Nachricht steht ein bestimmter endlicher Zeichenvorrat - ein Alphabet - zur Verfügung. Die einzelnen Zeichen dieses Alphabets weisen bestimmte Häufigkeiten bzw. Wahrscheinlichkeiten auf. Die statistischen Wahrscheinlichkeiten der einzelnen Zeichen, $P(i)$, summieren sich natürlich für das jeweilige Alphabet zu

$$\sum P_i = 1$$

Eine andere wichtige Eigenschaft von Zeichen ist, daß sie oft - wie bei den Buchstaben der natürlichen Sprache - in bestimmten kombinatorischen Formen auftreten oder gewisse Unverträglichkeiten aufweisen u. ä. m. .

Das heißt aber, daß die Entscheidungsstruktur der Nachrichtenquelle in diesen Fällen keine Unabhängigkeit der Zeichen zuläßt. Danach haben die einzelnen Zeichen nicht nur eine Wahrscheinlichkeit, sondern auch eine bedingte Wahrscheinlichkeit: die Wahrscheinlichkeit, daß bei diesem Entscheidungsprozeß das Zeichen i ausgewählt wird unter der Bedingung, daß das Zeichen j schon ausgewählt wurde, d.h., daß das letzte Zeichen j war. ($P(i | j)$). Die Wahrscheinlichkeiten und bedingten Wahrscheinlichkeiten der Zeichen bestimmen unter diesem Gesichtspunkt die statistische Struktur der Nachricht. ⁵⁴⁾

Da die Zeichen, "vermöge ihrer *Fähigkeit zur Auswahl*," einen Informationsgehalt besitzen, muß die Menge der Alternativen, d.h. das Alphabet, genau angegeben werden, bevor man den quantitativen Informationsgehalt von Nachrichten ermitteln kann. ⁵⁵⁾

9.2 Die statistische Konzeption der Informationsquantität.

Der Prozeß der Entstehung einer komplexen Nachricht vom Gesichtspunkt ihrer statistischen Struktur aus gesehen läßt sich folgendermaßen beschreiben:

Wenn wir eine Quelle mit einem Zeichenvorrat von n Zeichen betrachten

$$n = \{Z_1, Z_2, \dots, Z_n\},$$

die die statistischen Wahrscheinlichkeiten P_1, P_2, \dots, P_n aufweisen, dann ist die Nachrichtenproduktion eine aus einem Selektionsprozeß der Quelle aus dem Zeichenrepertoire entstandene Zeichenfolge. Der Informationsgehalt I_i , der einem beliebigen Zeichen Z_i entspricht, wird auf jeden Fall eine Funktion seiner reziproken Wahrscheinlichkeit sein, d.h.

$$I_i = f \left(\frac{1}{P_i} \right)$$

Dieser funktionale Zusammenhang wird als der Logarithmus der reziproken Wahrscheinlichkeit identifiziert und expliziert:

$$I_i = \log \frac{1}{P_i} = - \log P_i$$

Der mittlere Informationsgehalt einer Folge von Zeichen im Bezug auf den Zeichenvorrat beträgt dann

$$H = - \sum_{i=1}^n P_i \log P_i$$

Diese Zahl nennen wir in Anlehnung an die von Boltzmann in der Thermodynamik abgeleitete Beziehung Entropie (manchmal auch Negentropie, da sie im Gegensatz zur physikalischen Formel ein negatives Vorzeichen hat).

Durch die Entropie können wir eine quantitative Messung des durchschnittlichen Informationsgehalts vornehmen. Als Maßeinheit dient die Information, - der Grad der Unbestimmtheit-, die durch die Selektion eines Zeichens aus einem Alphabet von zwei gleichwahrscheinlichen Zeichen vermittelt wird. Diese Einheit wird ein "bit" genannt (Abkürzung aus dem Englischen binary digit = Binärziffer). Die Wahl dieser Maßeinheit bestimmt die Basis des Logarithmus der bei der Entropieformel verwendet wird, nämlich der dyadische Logarithmus $\lg P_i$ (der Logarithmus zur Basis 2).

Die wichtigsten Eigenschaften der Entropieformel sind folgende:

- (1) Der Wert von H ist immer positiv für jede P_i ist $P_i \leq 1$ und $\lg P_i \leq 0$; wegen des Minuszeichens vor der Formel ist $H \geq 0$.
- (2) Wenn der Zeichenvorrat ein einziges Zeichen besitzt, dann ist

der Ausgang der Selektion ein sicheres Ereignis, d.h. es existiert überhaupt keine Alternative, dieses Zeichen hat die Wahrscheinlichkeit $P_s = 1$ und infolgedessen ist $H = 0$: es besteht keine Ungewißheit und daher keine Information.

- (3) Die Entropie H hat einen maximalen Wert bei einer gegebenen Zeichenmenge der Quelle n , wenn alle Wahrscheinlichkeiten P_i gleich sind, nämlich $P_i = \frac{1}{n}$. Dieser Wert ist

$$H_{\max} = \log n$$

Dieses Resultat läßt sich mathematisch beweisen, ⁵⁶⁾ ist aber auch ohne diese Beweisführung einsichtig, denn gleiche Wahrscheinlichkeiten der verschiedenen Elemente implizieren größte Ungewißheit bei einer Auswahl und d.h. maximalen mittleren Informationsgehalt.

- (4) Wenn eine komplexe Nachricht aus einer Sequenz von unabhängigen Selektionsereignissen besteht, dann ist der Informationsgehalt dieser komplexen Nachricht gleich der Summe der Beiträge des Informationsgehalts jedes einzelnen Zeichens.

Die Formel für die Entropie ist tatsächlich nur ein Maß für eine Komponente des Begriffs der Information; sie gibt lediglich, auf ein Repertoire von Ereignissen, Zuständen u. ä. m., d.h. auf einen Zeichenvorrat bezogen, die statistische Ungewißheit oder den innovativen Wert einer Zeichenquelle an.

Die Verwandtschaft zwischen dem Entropiebegriff in der Thermodynamik (besser noch in der statistischen Mechanik) und dem in der Informationstheorie ist tiefergehend als die oberflächliche weitgehende Übereinstimmung des mathematischen Ausdrucks. Die Entropie kann hier wie dort als Maß für die Ordnung, für die Organisation eines Systems angenommen werden. Je genauer unsere Kenntnisse über ein physikalisches System sind, um so geringer ist die Entropie. Wenn wir z.B. Gas in einem abgeschlossenen Behälter haben, dann können wir einige Eigenschaften des Systems wie Druck und Temperatur ermit-

keln; wir können aber nicht solche Faktoren wie Lage und Geschwindigkeit aller Moleküle des Systems genau bestimmen, obwohl sie letzten Endes die Temperatur und den Druck bestimmen. Entropie kann also in diesem Fall als Maß für die Ungenauigkeit unseres Wissens hinsichtlich der genauen Lage und Geschwindigkeit der einzelnen Moleküle angesehen werden. ⁵⁷⁾

“Wie der Informationsgehalt eines Systems ein Maß für den Grad der Ordnung ist, ist die Entropie eines Systems ein Maß für den Grad der Unordnung; und das eine ist einfach das Negative des anderen”. ⁵⁸⁾

Durch die Einsetzung des Minuszeichens vor die Entropieformel werden in der Informationstheorie die zwei Maße gleichgesetzt und der Informationsgehalt eines Zustandes, einer Ereignismenge, mittels der Entropie errechnet.

Der informationstheoretische Ordnungsbegriff eines Systems läßt sich anschaulich an folgenden Schemata darstellen. ⁵⁹⁾

Ein solches Schema schreibt man, indem man der erwarteten Zeichenfolge Z_1, Z_2, \dots, Z_n die jeweiligen Selektions- bzw. Eintrittswahrscheinlichkeiten P_1, P_2, \dots, P_n zuordnet. Wenn wir nun von einem Alphabet mit vier Zeichen ausgehen und die entsprechenden jeweiligen Wahrscheinlichkeiten variieren, erhalten wir:

(a) $Z_1 \quad Z_2 \quad Z_3 \quad Z_4$

1/4 1/4 1/4 1/4

$H_a = 2,00 \quad \text{bit}$

(b) $Z_1 \quad Z_2 \quad Z_3 \quad Z_4$

1/2 1/4 1/8 1/8

$H_b = 1,75 \quad \text{bit}$

(c) $Z_1 \quad Z_2 \quad Z_3 \quad Z_4$

3/4 1/8 1/8 0

$H_c = 1,06 \quad \text{bit}$

(d) $Z_1 \quad Z_2 \quad Z_3 \quad Z_4$

$7/8 \quad 1/8 \quad 0 \quad 0$

$H_d = 0,5436 \text{ bit}$

(e) $Z_1 \quad Z_2 \quad Z_3 \quad Z_4$

$1 \quad 0 \quad 0 \quad 0$

$H_e = 0 \text{ bit}$

Unter Zugrundelegung des informationstheoretischen Ordnungsbegriffs lassen sich diese Schemata leicht interpretieren, wobei unter Ordnung eines Systems der durch eine Wahrscheinlichkeitsverteilung der Realisierungsmöglichkeiten seiner Elemente - Zeichen - charakterisierte Zustand verstanden wird, womit auch die Beziehungen zwischen Ordnung und Innovation bzw. Bestimmtheit und Information beschrieben werden.

Schema (a) zeigt den Zustand maximaler Unordnung d.h. maximaler Unbestimmtheit und Schema (e) den Zustand maximaler Ordnung, d.h. absoluter Bestimmtheit an. Die Schemata (b), (c), (d) zeigen den Übergang vom Zustand maximaler Unbestimmtheit zum Zustand absoluter Bestimmtheit an. Entsprechend ist auch der Übergang der Größe des quantitativen Informationsbetrages von dem größten durchschnittlichen Informationsgehalt bei maximaler Unordnung zum kleinsten, nämlich 0, beim Zustand der maximalen Ordnung der Elemente eines Systems.

Wir haben an anderer Stelle darauf hingewiesen, daß der informationstheoretische Ansatz seine Geltung und Anwendungsmöglichkeit auch außerhalb des nachrichtentechnischen Bereichs beibehält, und daß der Entropiebegriff auch auf diese Bereiche übertragen werden kann.

“Es ist möglich, das Rechnungsschema, das der Physiker benutzt, um die ‘Entropie’, also Zustände der Ungeordnetheit, der Durchmischung - um es nicht nur abstrakt zu sagen - zahlenmäßig auszudrücken, von der physikalischen (genauer: thermodynamischen) Realität auf die ästhetische Realität als ein mehr oder weniger gegliedertes oder ge-

ordnetes System materialer Elemente aufgefaßt werden, und es bleibt im Verhältnis zum darauf angewendeten abstrakten Rechnungsschema völlig gleichgültig, ob man dabei von einem System körperlicher Partikel namens 'Gas', einem System verbaler Zeichen namens 'Text' oder einem System farbiger Flecken namens 'Bild' spricht". ⁶⁰⁾

9.3 Relative Entropie und Redundanz

Wie wir gesehen haben, ist die maximale Entropie einer Nachrichtenquelle (oder wie wir jetzt auch sagen können einer Häufigkeits- oder Wahrscheinlichkeitsverteilung) dann gegeben, wenn die Selektionswahrscheinlichkeiten der Zeichen eines Alphabets gleich sind:

$H_{\max} = \text{Idn}$. Es ist offensichtlich, daß jede Änderung in der gleichwahrscheinlichen Verteilung der Zeichenselektion eine Abnahme der Entropie verursacht. Mit dem Anwachsen der übersehbaren Ordnung der Elemente eines Systems sinkt der Informationsgehalt, d.h. die Unbestimmtheit einer Nachricht. Die relative Unbestimmtheit einer Informationsquelle

$$H_{\text{rel}} = \frac{H}{H_{\max}}$$

gibt dann den Grad der Unbestimmtheit des realisierten Zustands, H , bezogen auf den potentiell maximalen Informationsgehalt dieser Quelle, H_{\max} , wider und heißt relative Entropie, H_{rel} . Aus der relativen Entropie leitet man durch

$$R = 1 - H_{\text{rel}} = 1 - \frac{H}{H_{\max}}$$

die Redundanz der Verteilung, die den Verlust an Unbestimmtheit mißt, der durch die Abweichung von der gleichwahrscheinlichen Selektionsverteilung der Zeichen hervorgerufen wird. ⁶¹⁾

Wenn die Zeichen eines Repertoires gleiche Eintrittswahrscheinlichkeiten haben und unabhängig voneinander sind, dann ist die Redundanz gleich Null, je größer die statistischen Einschränkungen

sind, desto höher ist die Redundanz. ⁶²⁾

Diesen Tatbestand können wir an Hand der fünf Schemata des letzten Abschnitts veranschaulichen.

So ist $H_{\max} = 2$ und

$H_{\text{rel}} (a) = 1$	$R (a) = 0$
$H_{\text{rel}} (b) = 0,875$	$R (b) = 0,125$
$H_{\text{rel}} (c) = 0,53$	Daraus folgt: $R (c) = 0,47$
$H_{\text{rel}} (d) = 0,2718$	$R (d) = 0,7282$
$H_{\text{rel}} (e) = 0$	$R (e) = 1$

“Die Ordnung wird . . . durch das statistische Maß für *Redundanz* bestimmt, bei der es sich im Prinzip um ein Maß dessen handelt, was innerhalb einer *Verteilung* der Elemente so stark determiniert ist, daß es vorhergesehen, identifiziert, erkannt werden kann, wie das mit jeder Ordnung der Fall ist”. ⁶³⁾

9.4 Informationsgehalt in Sequenzen abhängiger Zeichen

Die im letzten Abschnitt erwähnte Abweichung vom Zustand maximaler Unbestimmtheit kann auch darauf zurückzuführen sein, daß Zeichenfolgen bestehen, bei welchen die einzelnen Zeichenereignisse statistisch nicht unabhängig voneinander sind, d.h. die Eintrittswahrscheinlichkeit eines Zeichens an einer bestimmten Stelle auch davon abhängt, welche Zeichen vor ihm eingetreten sind. In natürlich-sprachlichen Kommunikationssystemen sind solche Zeichenfolgen weit verbreitet.

9.4.1 Markoff-Ketten

Bis jetzt haben wir Zeichenfolgen (Ereignisfolgen) betrachtet, deren

Zeichenereignisse unabhängig von dem Auftreten anderer Zeichen realisiert werden. Mit jeder Zeichenrealisation Z_i aus einem Ereignisraum A (Alphabet) ist eine Wahrscheinlichkeit P_i verbunden und die Wahrscheinlichkeit der Zeichenfolge setzt sich multiplikativ aus den einzelnen Wahrscheinlichkeiten P_i zusammen

$$P(Z_1, Z_2, \dots, Z_n) = P_1 \cdot P_2 \cdot \dots \cdot P_n$$

Von großer Bedeutung für die Informationstheorie ist die Tatsache, daß einzelne Zeichenereignisse nicht ganz unabhängig auftreten, sondern einer bestimmten Abhängigkeitsrelation unterworfen sind, so daß die Realisation eines Zeichens nicht mit einer bestimmten konstanten Wahrscheinlichkeit verbunden ist, sondern von dem unmittelbar vorhergehenden Zeichenereignis (oder von den unmittelbar vorhergehenden Ereignissen) abhängig ist.

Aus der Linguistik könnte man ganz einfache anschauliche Beispiele für solche Abhängigkeitsverhältnisse vorbringen, z.B. auf die Buchstabenfolge *sc* folgt in der deutschen Sprache fast immer nur (also mit großer Wahrscheinlichkeit) der Buchstabe *h*, oder auf die Wortfolge "mir geht" folgt sehr häufig (also mit verhältnismäßig großer Wahrscheinlichkeit) das Wort "es". Zu jedem Ereignispaar (Z_i, Z_j, Z_k) usw.) entspricht eine bedingte Wahrscheinlichkeit $P(Z_j | Z_i) = P_{ij}$, die Wahrscheinlichkeit, daß ein bestimmtes Ereignis Z_j eintritt unter der Bedingung, daß das Ereignis Z_i schon eingetreten ist.

Wenn die Abhängigkeit zwischen den Ereignissen über das unmittelbar vorhergehende Ereignis hinausgeht, sprechen wir von einem komplexen (diskreten) Markoff-Prozeß, sonst von einer Markoff-Kette. In jedem Markoff-Prozeß muß sich aber die Abhängigkeit über eine *endliche* Anzahl von Ereignissen erstrecken, d.h. die Anzahl der Zeichen, die die Wahrscheinlichkeit für das Auftreten eines bestimmten Zeichens beeinflussen, muß endlich sein.

Charakteristisch für die Markoff-Ketten ist also, daß mit der Realisierung der i -ten Selektion das *Zufallsgesetz* der $(i + 1)$ -ten Selektion eindeutig festgelegt ist, unabhängig davon, wie die $(i - 1)$ -te Selektion ausgefallen ist.

Bei den Markoff-Ketten wird eine etwas modifizierte Terminologie verwendet: so spricht man anstatt von Elementarmodalitäten und -ereignissen bzw. Zeichenereignissen auch von *Zuständen* des Systems und bei P_{ij} spricht man dann von der *Übergangswahrscheinlichkeit* vom i -ten in den j -ten Zustand.

Die bisherigen Ausführungen betreffen alle Prozeßzustände einer Kette, außer das Ausgangsglied. Wenn wir den Anfangszustand mit Z_{i0} bezeichnen, dann gibt P_{i0} die Wahrscheinlichkeit dafür an, daß die Kette mit dem Zustand Z_i beginnt. Damit läßt sich eine Markoff-Kette vollständig beschreiben: Die Wahrscheinlichkeit dafür, daß auf den Anfangszustand Z_i der Zustand Z_j folgt ist

$$P(Z_i) \cdot P(Z_j | Z_i) = p_i \cdot p_{ij}$$

und dafür, daß auf den Zustand Z_j der Zustand Z_k folgt

$$P(Z_i) \cdot P(Z_j | Z_i) \cdot P(Z_k | Z_j) = p_i \cdot p_{ij} \cdot p_{jk}$$

Wir können nun verallgemeinern, so daß für eine endliche Kette mit n Zuständen (wir indizieren nunmehr das eine Subskript i , wobei $i0$ den Anfangszustand darstellt) die Wahrscheinlichkeiten erhalten

$$P(Z_{i0}, Z_{i1}, \dots, Z_{in})_k = P(kn) \\ = p_{i0} \cdot p_{i0i1} \cdot p_{i1i2} \cdot \dots \cdot p_{in-2in-1} \cdot p_{in-1in}$$

Es wird dabei unterstellt, daß die Selektionen (Versuche) so einheitlich vorgenommen werden, daß die Indizierung der Übergangsschritte als Zeitparameter betrachtet wird.

Die Übergangswahrscheinlichkeiten werden in einer *Übergangsmatrix* zusammengestellt, die die Schritte und die entsprechenden Wahrscheinlichkeiten anschaulich darstellt und die Markoff-Kette vollständig beschreibt. Sie sieht folgendermaßen aus:

nach von					
	Z_1	$Z_2 \dots$	$Z_j \dots$	Z_n	
Z_1	p_{11}	$p_{12} \dots$	$p_{1j} \dots$	p_{1n}	$= P = [p_{ij}]$ $(i, j = 1, 2, \dots, n)$
Z_2	p_{21}	$p_{22} \dots$	$p_{2j} \dots$	p_{2n}	
\vdots	\vdots	\vdots	\vdots	\vdots	
\vdots	\vdots	\vdots	\vdots	\vdots	
Z_i	p_{i1}	$p_{i2} \dots$	$p_{ij} \dots$	p_{in}	
\vdots	\vdots	\vdots	\vdots	\vdots	
Z_n	p_{n1}	$p_{n2} \dots$	$p_{nj} \dots$	p_{nn}	

In der i-ten Zeile sind die Übergangswahrscheinlichkeiten vom Zustand i zu den Zuständen j und in der j-ten Spalte die Wahrscheinlichkeiten für die Übergänge aus den Zuständen j in den Zustand i aufgetragen.

Wenn wir nun ein einfaches Beispiel einer solchen Übergangsmatrix betrachten mit den drei Zuständen Z_1, Z_2, Z_3 und folgenden Übergangswahrscheinlichkeiten,

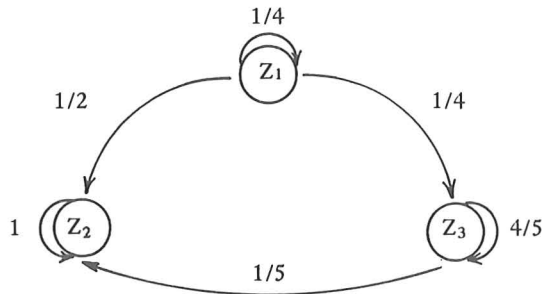
nach von				
	Z_1	Z_2	Z_3	
Z_1	1/4	1/2	1/4	$= P = [p_{ij}]$ $(i, j = 1, 2, 3)$
Z_2	0	1	0	
Z_3	0	1/5	4/5	

dann können wir diese Matrix folgendermaßen interpretieren:
ist das System im Zustand Z_1 , so bleibt es mit einer Wahrscheinlich-

keit von $1/4$ in dem gleichen Zustand, es geht mit einer Wahrscheinlichkeit von $1/2$ in den Zustand Z_2 über und mit einer Wahrscheinlichkeit von $1/4$ in den Zustand Z_3 über.

Wenn sich das System im Zustand Z_2 befindet, so geht es in keinen anderen Zustand über. Das ist ein spezieller Fall und man spricht vom Zustand Z_2 von einem absorbierenden Rand oder absorbierenden Zustand. Gelangt das System in den Zustand Z_3 , so führt von Z_3 nach Z_1 kein Weg, es geht mit einer Wahrscheinlichkeit von $1/5$ in den Zustand Z_2 und mit $4/5$ in Z_3 über.

Diese Beschreibung kann man anhand eines Übergangsdiagramms veranschaulichen.



Eine komplexe Nachricht ist in der Informationstheorie ein stochastischer Prozeß: eine Zeichenfolge (Buchstaben, Wörter oder sonstige materielle Informationsträger), deren einzelne Zeichen ausschließlich aufgrund einer Wahrscheinlichkeit gewählt worden sind, die von dem einen, den zwei, oder mehr realisierten Zuständen abhängt, die unmittelbar vorangehen. ⁶⁴⁾

“Jeder Realisation (in welchem Gebiet auch immer) liegt eine Kette von *Entscheidungen* zugrunde; jede Realisation ist Ausdruck einer *Auswahlfunktion*, in der material-selektive und normative Momente zusammentreten.” ⁶⁵⁾

Hier und im weiteren werden wir uns nur mit stochastischen Folgen befassen, in welchen Wahrscheinlichkeitsrelationen zwischen unmittelbar benachbarten Zuständen (Zeichen) definiert sind: mit Mar-

koff-Ketten. ⁶⁶⁾

9.4.2 Bedingte Entropie

Wenn wir nun den Informationsgehalt einer Sequenz abhängiger Zeichenereignisse betrachten, dann gibt die Entropieformel, die in Abschnitt 2 aufgeführt wurde, offensichtlich nicht den Informationsgehalt einer Zeichensequenz wider, deren Zeichen bekannte Übergangswahrscheinlichkeiten besitzen. Die Unbestimmtheit, die eine Situation vor dem Eintritt des nächsten Zeichens aufweist, kann nach der schon verwendeten Entropieformel berechnet werden, wenn anstatt der absoluten Wahrscheinlichkeit, die wegen der Abhängigkeit des noch zu realisierenden vom zuletzt realisierten Zustand nunmehr relevanten Übergangswahrscheinlichkeiten eingesetzt werden. Daraus läßt sich der Informationsgehalt eines Zustandes für eine fixierte Vorbedingung durch die Formel

$$H_j^i = - \sum_j P(j | i) \lg P(j | i) = - \sum_j P_{ij} \lg P_{ij}$$

ableiten, wobei i eine fixierte Zeichenrealisation und $j = 1, \dots, n$ ist.

Der Wert von H_j^i wird natürlich für die gleiche Zeichensequenz mit der jeweiligen Fixierung der Vorbedingung i variieren. So wird z.B. der Wert von H_j^i gleich Null sein, wenn das fixierte Zeichenereignis i der Buchstabe q ist, denn die Situation nach der Realisation von q ist keineswegs unbestimmt, nach q folgt immer u . Bei anderen Zuständen i wird H_j^i einen höheren Wert haben. Die durchschnittliche Unbestimmtheit in allen Übergangssituationen, bei welchen jeweils ein bekanntes Zeichenereignis fixiert ist, nennt man bedingte Information oder bedingte Entropie.

Da wir auch nicht den Anfangszustand einer Kette kennen, müssen wir die Eintrittswahrscheinlichkeit des Anfangszustandes, d.h. die Wahrscheinlichkeit dafür, daß eine Kette mit einem bestimmten Zeichen beginnt, berücksichtigen. Offensichtlich handelt es sich dabei

um eine absolute Wahrscheinlichkeit, nämlich die für die unabhängige Selektion eines Zeichens aus dem gesamten Reservoir, dem Zeichenvorrat.

Die bedingte Entropie, H_2 , läßt sich damit durch die Formel die wir hier herleiten, ermitteln.

$$\begin{aligned}
 H_2 &= - P(z_1) \sum_j P(z_j | z_1) \lg P(z_j | z_1) \dots - P(z_i) \sum_j P(z_j | z_i) \lg P(z_j | z_i) \\
 &\quad - P(z_n) \sum_j P(z_j | z_n) \lg P(z_j | z_n) \\
 &= - P_1 \sum_j P_{1j} \lg P_{1j} - P_i \sum_j P_{ij} \lg P_{ij} \dots - P_n \sum_j P_{nj} \lg P_{nj} \\
 &= - \sum_i \sum_j P_i P_{ij} \lg P_{ij} = \sum_i \sum_j P_i H_j^i
 \end{aligned}$$

Es ist manchmal einfacher und für manche Zwecke adäquater und anschaulicher die Berechnung der bedingten Entropie in einer Weise vorzunehmen, die etwas von der hier dargestellten Art abweicht. Wenn wir die Wahrscheinlichkeit aller Zeichenpaare $P(j,i)$ kennen, dann kann man die Erweiterung der Entropieformel für zweidimensionale Zeichenereignisse leicht vornehmen. Die so definierte Entropie wird dann den Informationsgehalt eines zweidimensionalen Zeichenereignisses - eines Zeichenpaares - messen. Diese zweidimensionale Entropie bezeichnen wir H^2 und berechnen sie nach der Formel:

$$H^2 = - \sum_i \sum_j P(j,i) \lg P(j,i)$$

Diese Formel gibt die durchschnittliche Information aller Digrammen oder Dyaden wider.

Aus dieser Formel und unter Zugrundelegung der bekannten Beziehung

$$P(j \wedge i) = P(j | i) P_i = P(j,i) = P_{ij} P_i$$

können wir umformen und erhalten

$$\begin{aligned}
 H^2 &= - \sum_i \sum_j P_{ij} P_i (\ln P_{ij} + \ln P_i) \\
 &= - \sum_i P_i \ln P_i \left(\sum_j P_i \sum_j P_{ij} \ln P_{ij} \right)
 \end{aligned}$$

Da der Ausdruck $\sum_j P_{ij} = 1$ ist, ergibt sich

$$H^2 = H + H_2 \quad \text{und} \quad H_2 = H^2 - H$$

Man kann also die bedingte Entropie auch als Differenz zwischen zweidimensionaler Verbundentropie und Entropie bei statistischer Unabhängigkeit der Zeichenereignisse errechnen.

Explication mathematischer Symbole

*Die Mathematiker sind eine Art
Franzosen; redet man zu ihnen, so
übersetzen sie es in ihre Sprache,
und dann ist es alsobald ganz etwas
anderes.*

J.W.v. Goethe

1. Das Summenzeichen (Σ)

Wenn wir die Summe $x_1 + x_2 + x_3 + x_4 + x_5$ haben, dann zeigt der Index die Stellung des einzelnen x-Wertes in dieser Reihe an. Die Addition soll von x_1 bis x_5 durchgeführt werden. Der Index i von x muß also von 1 bis 5 laufen.

(a)
$$\sum_{i=1}^5 x_i = x_1 + x_2 + x_3 + x_4 + x_5$$
 und allgemein

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_{n-1} + x_n$$

(b)
$$\sum_{i=1}^k f_i x_i = f_1 x_1 + f_2 x_2 + \dots + f_k x_k$$

(c)
$$\sum_{i=1}^n (x_i \pm y_i) = \sum_{i=1}^n x_i \pm \sum_{i=1}^n y_i$$

(d)
$$\sum_{i=1}^n (x_i \pm c) = \sum_{i=1}^n x_i \pm nc$$
 $c = \text{Konstante}$

(e)
$$\sum_{i=1}^n c x_i = c \sum_{i=1}^n x_i$$

2. Das Fakultätszeichen (!)

(a) $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$ (lies: 5 Fakultät)
 $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1) \cdot n$

(b) $\binom{6}{3} = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} = 20$ (lies: 6 über 3) oder

$$\binom{6}{3} = \frac{6!}{3! (6-3)!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 3} = 20$$

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot k} = \frac{n!}{k! (n-k)!}$$

(c) $\binom{n}{k} = \binom{n}{n-k}$

$$\binom{n}{0} = 1 \quad ; \quad \binom{n}{n} = 1$$

Anmerkungen

- 1) Zu diesem Komplex vgl. auch meine Ausführungen in meinem Aufsatz *Qualität, Quantität und Meßbarkeit*; in: S. Jäger (Hrsg.) *Linguistik und Statistik, Schriften zur Linguistik, Band 6*, Braunschweig 1972.
- 2) Daß die wissenschaftstheoretische Ungeklärtheit der Implikationen solcher Begriffe wie "Quantität", "quantitativ" und "Quantifizierbarkeit" nicht ohne Folgen für die wissenschaftliche Argumentation in der Linguistik bleibt, beweisen einige Ausführungen gerade solcher Autoren, die sich theoretisch am erfolgreichsten mit den Möglichkeiten und Grenzen quantitativer Methoden in der Linguistik beschäftigt haben.

Von der Opposition *langue* als abstraktes System versus *parole* als konkrete Aktualisierung ausgehend, schreibt Klaus Heger, zu den Methoden und Möglichkeiten einer quantitativen Linguistik; in: *Zeitschrift für Romanische Philologie*, Bd 20, 1964, S. 330: "*Bedenkt man weiterhin, daß zwar Quantitäten durch Abstraktion isolierbar - hierauf baut jede erfolgreiche quantitative Linguistik auf -, Abstraktionen aber nicht quantifizierbar sind*, so läßt sich daraus unschwer der Schluß ziehen, daß eine ausschließlich mit der *langue* im Sinne eines abstrakten Systems verstandenen Sprache befaßte Sprachwissenschaft mit quantitativen Methoden wenig anfangen kann . . ." (meine Hervorhebungen mit Ausnahme von *langue*. P.N.). Hier handelt es sich offensichtlich um verschiedene Momente des Abstraktionsbegriffs, die in ihren unterschiedlichen Ausprägungen unterschiedlicher Behandlungen bedürfen. Daß "Abstraktionen . . . nicht quantifizierbar" seien, kann zweierlei bedeuten, nämlich entweder den Prozeß, der zu abstrakten Gebilden - welcher Abstraktionsstufe auch immer - führt, oder das Ergebnis dieses Abstraktionsprozesses (worauf die Schlußfolgerung im Hinblick auf das abstrakte System hinweist). Im ersten Fall würde die Feststellung der Nicht-Quantifizierbarkeit in der hier

anvisierten allgemeinen Form zutreffen, die Schlußfolgerung auf das abstrakte System der langue aber irrelevant sein. Im zweiten Fall, nämlich Abstraktion als Produkt des Abstraktionsprozesses, würde die Feststellung der Nicht-Quantifizierbarkeit in dieser allgemeinen Form nicht zutreffen, denn es hängt davon ab, ob diese Abstraktionsprodukte die Bedingungen verschiedener Quantifizierungsniveaus erfüllen oder nicht. Die Schlußfolgerung auf das abstrakte System ist dann in dieser Form nicht mehr möglich. Vgl. hierzu auch Kap. C dieser Arbeit.

- 3) Vgl. z.B. folgende Ausführungen von M.A. Girshick in einer Rezension des Buches von M.J. Moroney: *Facts from Figures*, Baltimore, Maryland 1951: "All branches of statistics . . . deal with the same basic problem , namely, the problem of decision making in the face of uncertainty . . . not facts from figures but rather decisions from observations should become the main emphasis in elementary statistical instructions". In; *Journal of the American Statistical Association*, Bd. 48 (1953), S. 646.
- 4) C. Wright Mills: *Kritik der soziologischen Denkweise*, Soziologische Texte, Band 8, Neuwied/Rhein und Berlin-Spandau 1963, S. 270.
- 5) Als Beispiel für manche Gefahren, die aus der Partikularisierung eines komplexen Phänomens zwecks Erreichung der Manipulierbarkeit mittels statistischer Methoden, mögen Stiluntersuchungen in der statistischen Linguistik dienen. Die Häufigkeitsverteilungen inklusiv Rangaussagen und Verhältniszahlen von Types und Tokens u.ä.m. werden, oft (natürlich nicht bei denjenigen, die die Art des Vorgehens, wie wir sie beschrieben haben, zu Grunde legen) als eine hinreichende Stilcharakterisierung angesehen, was natürlich keineswegs die volle Wahrheit ist, abgesehen davon, daß man auch noch darüber diskutieren könnte, ob es überhaupt der wichtigste Teil dieser Wahrheit sei. Denn das gleiche "quantitative" Ergebnis auf dieser Ebene kann nicht nur verschiedene Ursachen haben, sondern darüberhinaus auch

noch verschiedene stilistische Ausprägungen charakterisieren, wobei hier unter Stil eine komplexe Eigenschaft verstanden wird, die vom Forscher oder vom Rezipienten zur Klassifikation unterschiedlicher textlicher Erzeugnisse - unter diesem Gesichtspunkt - zu Grunde gelegt wird. (Es ist eine Frage an die Theorie, diese komplexe Eigenschaft zu analysieren und ihre Elemente und die zwischen ihnen bestehenden Beziehungen anzugeben). Wenn man andererseits unter Stil die aus solchen Häufigkeitsverteilungen abgeleiteten, abzuleitenden oder ableitbaren Beziehungen oder Bestimmungen versteht, dann verkürzt man natürlich das Problem auf die unzulänglichen Dimensionen eines im wissenschaftlichen hic et nunc "quantitativ" erschließbaren Gegenstandes, daher auch Schein-Gegenstandes.

- 6) J. Pfanzagl:
Allgemeine Methodenlehre der Statistik. Bd. I: Elementare Methoden unter besonderer Berücksichtigung der Anwendungen in den Wirtschafts- und Sozialwissenschaften
1. Aufl. Berlin 1960, 3. Aufl. Berlin 1966, S. 14.
- 7) M. Hengst:
Einführung in die Mathematische Statistik und ihre Anwendung, Mannheim, Wien, Zürich 1967, S. 99.
- 8) wobei das Unterscheidungskriterium, d.h. welche Wörter als verschieden betrachtet werden, der theoretischen Linguistik überlassen werden muß.
- 9) Vgl. hierzu u.a. auch R.M. Frumkina: The Application of Statistical Methods in Linguistic Research, in: D.G. Hays and D.V. Mohr (vom Russischen übersetzt) Exact Methods in Linguistic Research, Berkeley and Los Angeles 1963, S. 97f.
- 10) P. Guiraud z.B. hat einen sehr einfachen Index vorgeschlagen, nämlich $\frac{V_1}{V}$, d.h. den Anteilswert der Wörter, die in einem Text einmal vorkommen (Häufigkeit 1) an der gesamten Wortliste. In unserer Schreibweise also

$$\frac{f_i}{k} = \frac{f_i}{N}$$

$$\sum_{i=1}^k f_i$$

Problèmes et Méthodes de la statistique linguistique, Paris 1960, S. 89.

11) G.U.Yule:

The statistical Study of Literary Vocabulary, Cambridge 1944, S. 55. Die Charakteristik K von Yule hat folgende Gestalt

$$K = 10^4 \left(\frac{m_2}{m_1^2} - \frac{1}{m_1} \right)$$

Der Faktor 10^4 wurde zur Vermeidung sehr kleiner Zahlen eingeführt. m_1 und m_2 sind das erste und zweite gewöhnliche oder Anfangsmoment der Verteilung mit

$$m_1 = \frac{1}{N} \sum_{i=1}^k f_i x_i \quad \text{und} \quad m_2 = \frac{1}{N} \sum_{i=1}^k f_i x_i^2.$$

Für große Stichproben kann man den zweiten Klammerausdruck vernachlässigen, so daß die Charakteristik die Form enthält:

$$K^* = \frac{m_2}{m_1^2} \quad \text{bzw.} \quad K^* = \frac{\sum_{i=1}^k f_i x_i^2}{\left(\sum_{i=1}^k f_i x_i \right)^2}.$$

12) Siehe u.a. G. Herdan: The Advanced Theory of Language as Choice and Chance, Berlin, Heidelberg, New York 1966, S. 101 ff.

- 13) In einem späteren Teil werden Fragen der Stichprobentheorie erörtert.
- 14) Siehe hierzu C. Muller: *Initiation à la statistique linguistique*, Paris 1968, vor allem sein Diagramm auf Seite 101 und G. Herdan: *The Advanced Theory of Language as Choice and Chance*, S. 96 ff und 102 ff.
- 15) R.M. Frumkina (Über das sogenannte "Zipfsche Gesetz"; in: U. Engel (Hrsg.) *Forschungsberichte des Instituts für deutsche Sprache*, Band 4, Mai 1970) spricht in diesem Zusammenhang von "Statistischer Struktur des Textes", wobei unter diesem Begriff "die Beziehung zwischen der Zahl verschiedener Wörter in einem gegebenen Text (d. i. dem Umfang des Wortbestandes) und der Häufigkeit Ihrer Wiederholung im Text zu verstehen" ist. (S. 118).

Mir scheint, daß diese Definition eine unnötige Einschränkung des Ausdrucks "statistische Struktur eines Textes" auf die o.e. Beziehung bedeutet. Wie soll man dann andere statistisch ermittelte oder zu ermittelnde Strukturen eines Textes bezeichnen? Die statistische Struktur eines Textes ist ein viel komplexeres und umfassenderes statistisch-linguistisches Beziehungssystem als die o.e. singuläre Beziehung. Die auf die Umgebung des Zipfschen Gesetzes zugeordnete - verwiesene - statistisch-linguistische Strukturdefinition bildet daher eine implizite begriffliche Schranke bei der Ausgestaltung der strukturellen Komponenten eines Textes in ihrer statistisch-linguistischen Mehrdimensionalität.

Frumkina selbst hat in einem anderen Artikel (*The Application of statistical Methods in Linguistic Research*) dem partikularen Charakter der o.e. Beziehung für die allgemeine Definition des Begriffs "statistische Struktur eines Textes" Rechnung getragen, indem sie geschrieben hat: "... we describe the 'Structure' of the text *from the standpoint of word repetition*, showing how many words have frequency 1, 2, etc. We are not dealing here with individual properties of words, and we extract only that

property which is important for characterizing text *from a particular standpoint*" (S. 97 f; meine Hervorhebungen P.N.).

- 16) G. Herdan:
About some controversial results of the quantitative method
in linguistics; in: Zeitschrift für Romanische Philologie, Band 85,
1969, S. 378.
- 17) S.S. Stevens:
Measurement, Statistics and the Schemapiric View; in: Science,
Vol. 161 (1968), wiederabgedruckt in und zitiert nach B. Lieber-
man (ed.), Contemporary Problems in Statistics. A Book of
Readings for the Behavioral Sciences, New York, London,
Toronto 1971, S. 101.
- 18) W. Leinfellner:
Einführung in die Erkenntnis- und Wissenschaftstheorie, 2. erw.
Auflage, Mannheim 1967, S. 132 f. Vgl. auch S. 59ff.
- 19) Vgl. J. Ritsert/E. Becker:
Grundzüge sozialwissenschaftlich-statistischer Argumentation,
Opladen 1971, S. 51.
- 20) B.S. Philips:
Empirische Sozialforschung, - Strategie und Taktik, Wien -
New York 1970, S. 223.
- 21) I.I. Revzin:
Models of Language, London 1966, S. 60.
- 22) Vgl. zu dem hier behandelten Komplex auch P. Nikitopoulos:
Qualität, Quantität und Meßbarkeit, a.a.O.
- 23) G.E. Peterson und Frank Harary:
Foundations of Phonemic Theory; in: R. Jakobson (ed.),
Structure of Language and its mathematical Aspects, Proceedings
of Symposia in Applied Mathematics, Vol. XII, Providence, Rhode
Island, 1961, S. 156.
- 24) Vgl. auch P. Nikitopoulos: Qualität, Quantität und Meßbarkeit,
a.a.O.

- 25) W.L. Hays:
Statistics, London, New York, Sydney, Toronto 1969, S. 72.
- 26) Zu einer kritischen Durchleuchtung der hauptsächlich auf Stevens zurückgehenden Adäquatheitsüberlegungen hinsichtlich der verschiedenen Skalenarten siehe: B. Ellis: Basic Concepts of Measurement, Cambridge 1968, vor allem den Abschnitt The appropriateness of statistics, S. 68ff. In diesem Abschnitt führt Ellis einen neuen Ansatz zur Überprüfung der Adäquatheit von Mittelwerten in Bezug auf die verschiedenen Skalenarten aus. Dieser Ansatz ist zwar sehr anschaulich, obwohl er in manchen Beziehungen unvollständig ist, wir verzichten aber auf seine Darstellung, da er außerhalb des intendierten Rahmens dieser Ausführungen liegt.
- 27) S.S. Stevens:
On the Theory of Scales of Measurement; in: Science, Vol. 103, (1946), wiederabgedruckt in und zitiert nach B. Lieberman (ed.) Contemporary Problems in Statistics, S. 4.
- 28) "If there should exist any data incompatible with the axioms of level of measurement selected, these data constitute error, by definition", C.H. Coombs: Theory and Methods of Social Measurement, in: Festinger/Katz (Eds.) : Research Methods in the Behavioral Sciences, New York 1965, S. 486.
- 29) G. Gäfgen:
Theorie der wirtschaftlichen Entscheidung. Untersuchungen zur Logik und ökonomischen Bedeutung des rationalen Handelns, 2. durchg. und erw. Auflage, Tübingen 1968, S. 141.
- 30) C.G. Osgood, G.J. Suci und P.H. Tannenbaum:
The Measurement of Meaning, Urbana 1957. Siehe dort auch für weitere Einzelheiten.
- 31) C.E. Osgood u.a.:
The Measurement of Meaning, S. 20. Zitiert nach der deutschen Übersetzung dieses Abschnitts, in: A.V. Cicourel: Methode und Messung in der Soziologie, Frankfurt/M. 1970, S. 259.

- 32) J.B. Carroll:
Language and Thought, Englewood Cliffs, New Jersey 1964,
S. 104.
- 33) Vgl. W.L. Hays:
Statistics, S. 75. Daher betont A.V. Cicourel, a.a.O. S. 20:
"So erfordern interpretierte axiomatische Systeme, daß eine
Entsprechung zwischen den Elementen, Beziehungen und
Operationen der betreffenden mathematischen Systeme be-
wiesen wird".
- 34) W.L. Hays:
Statistics, S. 74.
- 35) Vgl. zu diesem Beispiel K.R. Hammond und J.E. Householder:
Introduction to the statistical Method, Foundations and Use
in the Behavioral Sciences, New York 1967, S. 40f.
- 36) P. Nikitopoulos:
Qualität, Quantität und Meßbarkeit. Dort auch andere generelle
Probleme der Meßbarkeit im sozialwissenschaftlichen Objekt-
bereich.
- 37) Vgl. auch u.a. J. Habermas: Zur Logik der Sozialwissenschaften;
in: Philosophische Rundschau, Beiheft 5, Tübingen 1967 und
vom gleichen Autor die Einleitung "Einige Schwierigkeiten beim
Versuch, Theorie und Praxis, zu vermitteln" zur Neuausgabe von
4., durchg., erw. und neu ingl. Auflage, Frankfurt/M 1971. Dort
schreibt er: "Alle auf das Sprachspiel physikalischen Messens zu-
rückführbaren Operationen (auch an Meßgeräten, die nur mit
Hilfe komplizierter Theorien gebaut werden können) lassen sich
der sinnlichen Wahrnehmung ('Beobachtung') und einer Ding-Er-
eignis-Sprache zuordnen, in der Beobachtungen deskriptiv ausge-
drückt werden können. Hingegen fehlt ein entsprechendes System
von Grundoperationen des Messens, das wir in analoger Weise
dem in Zeichenbeobachtungen fundierten Sinnverstehen sowie
einer Person-Äußerungssprache zuordnen könnten, in der ver-
standene Äußerungen deskriptiv ausgedrückt werden. Wir be-

helfen uns mit hermeneutisch disziplinierten Deutungen, d.h. wir bedienen uns der Hermeneutik anstelle eines Meßverfahrens, aber sie ist keines". (S. 18). Er vermutet, daß eine Theorie, die die kommunikative Kompetenz erklären würde, "eine geregelte Umformung kommunikativer Erfahrungen in Daten erlauben" würde, und zwar "ähnlich wie . . . die Transformationsgrammatik Untersuchungen der Psycholinguistik über den kindlichen Spracherwerb der Konstruktion von Meßverfahren eine normative Grundlage" bietet. (S. 18-19).

- 38) M. Hengst:
Einführung in die Mathematische Statistik und ihre Anwendungen, S. 16.
- 39) G. Menges:
Grundriß der Statistik, Teil 1: Theorie, Köln und Opladen 1968, S. 28.
- 40) A.N. Kolmogoroff:
Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin 1933; hier zitiert nach P. Neurath: Statistik für Sozialwissenschaftler. Eine Einführung in das statistische Denken, Stuttgart 1966, S. 51.
- 41) L.J. Savage:
Foundations of Statistics, New York 1954, S.3.
- 42) Die Bezeichnungen und Abgrenzungen für die nicht elementaren Ereignisse des Mengenkörpers K (E) variieren sehr, wobei die Logik und Notwendigkeit der verschiedenen Einteilungen nicht immer klar sind und sehr oft Verwirrung stiften. Wir führen hier einige Beispiele auf:
M.Hengst (Einführung in die Mathematische Statistik und ihre Anwendung) definiert ein Ereignis als "zusammengesetzt, wenn es als Summe mindestens zweier Ereignisse A und B dargestellt werden kann, die beide von C verschieden sind", (S. 31) d.h. $A \cup B = C$ mit $C \neq B$. W.L. Hays (Statistics) definiert: "any event that is the intersection of two or more events is a joint event" (S. 108), d.h. $A \cap B = C$. Bei T.Yamane (Statistics:

An Introductory Analysis, 2nd ed., New York, Evanston & London-Tokyo 1964) heißt es: "An event that can be decomposed into simple (d.h. elementary P.N.) events is called a *compound event*" (S. 79), und D.Hérault (Éléments de théorie moderne des probabilités, Cours du Centre de Linguistique Quantitative 1, Paris 1967) findet, "il est préférable de définir les événements comme étant des sous-ensembles quelconques, ne contenant pas forcément un seul élément" (S. 69).

- 43) G. Menges:
Grundriß der Statistik, S. 109.
- 44) Vgl. u.a. G.Menges:
Grundriß der Statistik, S. 177f.
- 44) Vgl. auch J.Riordan:
An Introduction to Combinatorial Analysis, 4th print., New York, London, Sydney 1967, S. 90f.
- 46) Vgl. H.Margenau and G.M. Murphy:
The Mathematics of Physics and Chemistry, New York 1943, speziell chapter 12.
- 47) Vgl. u.a. auch W.Feller:
An Introduction to Probability Theory and Its Applications, Vol I, 3rd ed. New York, London, Sydney 1957, S. 40f, H. Meschkowski: Wahrscheinlichkeitsrechnung, Mannheim, Zürich 1968, S. 30ff, G.Menges: Grundriß der Statistik, S. 128ff, J.Riordan: An Introduction to Combinatorial Analysis, ch. 5.
- 48) E. Whittaker:
From Euclid to Eddington, Cambridge 1949.
- 49) An dieser Stelle mag der Hinweis auf zwei Arbeiten genügen, die die in diesen Abschnitten behandelten statistischen Lehren auf dem Gebiet der Linguistik anwenden.
In der einen Arbeit von P.Suppès: Probabilistic Grammars for Natural Languages, Institute for Mathematical Studies in the Social Sciences, Stanford University, California, Psychology Series, Technical Report No. 154, 1970 (vervielfältigtes Manus-

kript) wird davon ausgegangen, daß "objective probabilistic criteria of a standard scientific sort may be used to select a grammar" (S. 1) . Die gegenwärtige Haupttendenz in der Linguistik ist nach Suppes "to look at the deviant cases and not to concentrate on a quantitative account of that part of a corpus that can be analyzed in relatively simple terms"; es ist infolgedessen "one objective of a probabilistic grammar . . . to account for a high percentage of a corpus with relatively simple grammar and to isolate the deviant cases that need additional analysis and explanation" (S. 5). Was darüber hinaus aber wichtig ist, ist "that the objective of the probabilistic model is not just to give an account of the finite corpus of spoken speech or written text used as a basis for estimating the parameters of the model, but to use the finite corpus as a sample to infer parameter values for a larger, potentially infinite 'population' in the standard probabilistic fashion". (S. 5)

In der Arbeit von G.Herdan: *The Calculus of Linguistik Observations*, 'S-Gravenhage 1962, finden sich einige Anwendungen der nicht-klassischen Wahrscheinlichkeitsmodelle auf linguistische Fragestellungen, insbesondere in den Abschnitten XVII und XVIII. Im Rahmen dieser Arbeit ist es aber nicht möglich, auf solche spezielle und umfangreiche Arbeiten einzugehen; es würde außerdem den Intentionen dieser Arbeit, die relevanten statistischen Methoden darzulegen, widersprechen.

- 50) Vgl. G. Klaus (Hrsg.):
Wörterbuch der Kybernetik, Berlin 1968, S. 602.
- 51) Hier wird ersichtlich, warum die Festlegung, d.h. die Einengung des Begriffs der statistischen Struktur ausschließlich auf die von der Wortwiederholung her charakteristischen Situation eines Textes als partikularen Aspekt der Mehrdimensionalität des Strukturbegriffs abgelehnt wurde. Vgl. Anmerkung 15) des Kapitels B.
- 52) M.Bense:
Zusammenfassende Grundlegung moderner Ästhetik; in: H.

- Kreuzer und R.Gunzenhäuser (Hrsg.) Mathematik und Dichtung, 2. durchg. Aufl., München 1967, S. 326.
- 53) N.Wiener: Kybernetik, Düsseldorf und Wien 1963, S. 31.
- 54) Vgl. E.V.Paducheva:
Information Theory and the Study of Language; in: D.G.Hays and D.V.Mohr (vom Russischen übersetzt). Exakt Methods in Linguistic Research, Berkeley and Los Angeles 1963, S. 122.
- 55) Vgl. u.a. C.Cherry:
Kommunikationsforschung- eine neue Wissenschaft, 2. erw. Aufl., Frankfurt/M 1967, S. 217f.
- 56) Zur mathematischen Beweisführung siehe u.a. R. Deutsch: System Analysis Techniques, Englewood Cliffs, N.J. 1969, S. 421f und A.M.Jaglom und I.M.Jaglom: Wahrscheinlichkeit und Information, 3. bericht. Aufl., Berlin 1967, S. 53ff und S. 327ff .
- 57) Vgl. C.E.Shannon and W.W.Weaver:
The Mathematical Theory of Communication, Urbana, Illinois 1949, S. 95f und E.V.Paducheva: Information Theory and the Study of Language, S. 128 und J.R.Pierce: Phänomene der Kommunikation. Informationstheorie - Nachrichtenübertragung - Kybernetik, Düsseldorf-Wien 1965, S. 36ff .
- 58) N.Wiener: Kybernetik, S. 31.
- 59) Vgl. F.v.Cube und W.Reichert:
Das Drama als Forschungsobjekt der Kybernetik, in: H. Kreuzer und R. Gunzenhäuser (Hrsg.) Mathematik und Dichtung, S. 333f und M. Bense: Einführung in die informationstheoretische Ästhetik. Grundlegung und Anwendung in der Texttheorie, Reinbek bei Hamburg 1969, S. 57f.
- 60) M.Bense:
Zusammenfassende Grundlegung moderner Ästhetik, S. 321f .
Bense (Einführung in die informationstheoretische Ästhetik) bezeichnet die diesen o.a. Schemata entsprechenden ästhetischen

Zustände als *chaogen* (Schema a), *konfigurativ* (Schemata b, c, d) und *strukturell* (Schema e) (S. 58), wobei "mit dem Anwachsen der übersehbaren Ordnung (der Elemente) eines ästhetischen Zustandes . . . seine Innovation, seine Originalität (sinkt). Der chaogene Zustand der gleichwahrscheinlichen Distribution der Elemente hat insofern selbstverständlich ein Maximum der Innovation oder Information, als aus ihm jeder andere strukturelle oder konfigurative ästhetische Zustand selektiert werden kann. Jedes Chaos ist reale Quelle, reales Repertoire möglicher Innovationen im Sinne von Kreationen. Das gehört essentiell zur Erklärung des Chaos." (S. 58 - 59).

- 61) Vgl. auch G.Ungeheuer:
Language in the light of information theory; in: Intern. Soc. Sci.J., Vol. XIX, No. 1, 1967. Manchmal wird R als relative Redundanz und $R_a = H_{\max} - H$ als absolute Redundanz bezeichnet; siehe hierzu u.a. H. Meschkowski: Wahrscheinlichkeitsrechnung, S. 204, K. Steinbuch und W.Rupprecht: Nachrichtentechnik. Eine einführende Darstellung, Berlin, Heidelberg, New York 1967, S. 515.
- 62) E.V.Paducheva:
Information Theory and the Study of Language, S. 131.
- 63) M.Bense:
Einführung in die informationstheoretische Ästhetik, S. 106 .
- 64) Vgl. C.Cherry:
Kommunikationsforschung - eine neue Wissenschaft, S. 60 .
- 65) W.L.Fischer:
Zur Algebra der Texte, 2. Teil; in: Beiträge zur Literatur und Informationsverarbeitung, Heft 8, 1966, S. 56 .
- 66) Für eine umfassendere Darstellung der Markoff-Ketten siehe J.G.Kemeny and J.L.Snell: Finite Markov Chains, Princeton, N.J. 1960 .

Literaturverzeichnis

- Bense, M., Zusammenfassende Grundlegung moderner Ästhetik; in:
H. Kreuzer und R. Gunzenhäuser (Hrsg.) Mathematik und Dichtung, 2. durchges. Aufl., München 1967.
- Bense, M., Einführung in die informationstheoretische Ästhetik.
Grundlegung und Anwendung in der Texttheorie, Reinbek bei
Hamburg 1969.
- Carroll, J.B., Language and Thought, Englewood Cliffs, N.J. 1964.
- Cherry, C., Kommunikationsforschung - eine neue Wissenschaft,
2. erw. Aufl., (Frankfurt) 1967.
- Cicourel, A.V., Methode und Messung in der Soziologie, Frankfurt/
Main 1970.
- Coombs, C.H., Theory and Methods of Social Measurement; in:
Festinger/Katz (eds.) Research Methods in the Behavioral
Sciences, New York 1965.
- Cube, F. von und W. Reichert, Das Drama als Forschungsobjekt der
Kybernetik; in: H. Kreuzer und R. Gunzenhäuser (Hrsg.) Mathematik und Dichtung, 2. durchges. Auflage, München 1967.
- Deutsch, R., System Analysis Techniques, Englewood Cliffs, N.J.
1949.
- Ellis, B., Basic Concepts of Measurement, Cambridge 1968.
- Feller, W., An Introduction to Probability Theory and Its Applications, Vol. I, 3rd ed., New York-London-Sydney 1957.
- Fischer, W.L., Zur Algebra der Texte, 2. Teil; in: Beiträge zur
Linguistik und Informationsverarbeitung, Heft 8, München
und Wien 1966.
- Frumkina, R.M., The Application of Statistical Methods in Linguistic Research; in: D.G. Hays and D.Y. Mohr (transl. and eds.) Exact Methods in Linguistic Research, Berkeley and Los Angeles 1963.

- Frumkina, R.M., Über das sogenannte "Zipfsche Gesetz"; in: U. Engel (Hrsg.) Forschungsberichte des Instituts für deutsche Sprache, Bd. 4, Mannheim 1970.
- Gäfgen, G., Theorie der wirtschaftlichen Entscheidung. Untersuchungen zur Logik und ökonomischen Bedeutung des rationalen Handelns, 2. durchges. und erw. Aufl., Tübingen 1968.
- Girshick, M.A., Rezension des Buches von M.J. Moroney: Facts from Figures, Baltimore, Maryland 1951; in: Journal of the American Statistical Association, Bd. 48, 1953.
- Guiraud, P., Problèmes et Méthodes de la Statistique Linguistique, Paris 1960.
- Habermas, J., Zur Logik der Sozialwissenschaften; in: Philosophische Rundschau, Beiheft 5, Tübingen 1971.
- Habermas, J., Theorie und Praxis, 4. durchges., erw. und neu eingel. Aufl., Frankfurt/M. 1971.
- Hammond, K.R. and J.E. Householder, Introduction to the Statistical Method. Foundations and Use in the Behavioral Sciences, New York 1967.
- Haseloff, O.W. und H.-J. Hoffmann, Kleines Lehrbuch der Statistik für Naturwissenschaft und Technik, Psychologie, Sozialforschung und Wirtschaft, 3. Aufl., Berlin 1968.
- Hays, W.L., Statistics, London-New York-Sidney-Toronto 1969.
- Heger, K., Zu den Methoden und Möglichkeiten einer quantitativen Linguistik; in: Zeitschrift für Romanische Philologie, Bd. 20, 1964.
- Hengst, M., Einführung in die mathematische Statistik und ihre Anwendung, Mannheim-Wien-Zürich 1967.
- Hérault, D., Éléments de théorie moderne des probabilités, Cours du Centre de Linguistique quantitative 1, Paris 1967.
- Herdan, G., The Calculus of Linguistic Observations, 'S-Gravenhage 1962.

- Herdan, G., *The Advanced Theory of Language as Choice and Chance*, Berlin-Heidelberg-New York 1966.
- Herdan, G., About some controversial results of the quantitative method in linguistics; in: *Zeitschrift für Romanische Philologie*, Bd. 85, Heft 3/4, Tübingen 1969.
- Jaglom, A.M. und I.M. Jaglom, *Wahrscheinlichkeit und Information*, 3. bericht. Aufl., Berlin 1967.
- Kemeny, J.G. and J.L. Snell, *Finite Markov Chains*, Princeton, N.J. 1960.
- Klaus, G. (Hrsg.), *Wörterbuch der Kybernetik*, Berlin 1968.
- Leinfellner, W., *Einführung in die Erkenntnis- und Wissenschaftstheorie*, 2. erw. Aufl., Mannheim 1967.
- Lipschutz, S., *Theory and Problems of Set Theory and Related Topics*, New York-St. Louis-San Francisco-Sydney 1964.
- Lipschutz, S., *Theory and Problems of Probability*, New York-St. Louis-San Francisco-Sydney 1968.
- Margenau, H. and G.M. Murphy, *The Mathematics of Physics and Chemistry*, New York 1943.
- Menges, G., *Grundriß der Statistik. Teil 1: Theorie*, Köln und Opladen 1968.
- Menges, G., *Grundmodelle wirtschaftlicher Entscheidungen. Einführung in moderne Entscheidungstheorien unter besonderer Berücksichtigung volks- und betriebswirtschaftlicher Anwendungen*, Köln und Opladen 1969.
- Meschkowski, H., *Wahrscheinlichkeitsrechnung*, Mannheim-Zürich 1968.
- Mills, F.C., *Statistical Methods*, 3rd. ed., London 1955.
- Muller, C., *Essai de Statistique Lexicale*, Paris 1964.
- Muller, C., *Initiation à la statistique linguistique*, Paris 1968.
- Neurath, P., *Statistik für Sozialwissenschaftler. Eine Einführung in das statistische Denken*, Stuttgart 1966.

- Nikitopoulos, P., Qualität, Quantität und Meßbarkeit; in: S. Jäger (Hrsg.) Linguistik und Statistik, Schriften zur Linguistik, Bd. 6, Braunschweig 1972.
- Osgood, C.G., G.J. Suci and P.H. Tannenbaum, The Measurement of Meaning, Urbana, Illinois 1957.
- Paducheva, E.V., Information Theory and the Study of Language; in: D. G. Hays and D.V. Mohr (transl. and eds.) Exact Methods in Linguistic Research, Berkeley and Los Angeles 1967.
- Peterson, C.E. and F. Harary, Foundations of Phonemic Theory; in: R. Jakobson (ed.) Structure of Language and its Mathematical Aspects, Proceedings of Symposia in Applied Mathematics, Vol. XII, Providence, Rhode Island 1961.
- Phillips, B.S., Empirische Sozialforschung - Strategie und Taktik, Wien-New York 1970,
- Pierce, J.R., Phänomene der Kommunikation - Informationstheorie - Nachrichtenübertragung - Kybernetik, Düsseldorf-Wien 1965.
- Riordan, J., An Introduction to Combinatorial Analysis, 4th pr., New York-London-Sydney 1967.
- Ritsert, J. und E. Becker, Grundzüge sozialwissenschaftlich- statistischer Argumentation, Opladen 1971.
- Savage, L. J., Foundations of Statistics, New York 1954,
- Shannon, C.E. and W.W. Weaver, The Mathematical Theory of Communication, Urbana, Illinois 1949.
- Steinbuch, K. und W. Rupprecht, Nachrichtentechnik. Eine einführende Darstellung, Berlin-Heidelberg-New York 1967.
- Stevens, S.S., On the Theory of Scales of Measurement; in: Science, Vol. 103 (1946), wiederabgedruckt in und zitiert nach B. Lieberman (ed.) Contemporary Problems in Statistics. A Book of Readings for the Behavioral Sciences, New York-London-Toronto 1971.
- Stevens, S.S., Measurement, Statistics and the Schemapiric View, in: Science, Vol 161 (1968), wiederabgedruckt in und zitiert

- nach B. Lieberman (ed.) Contemporary Problems in Statistics, A Book of Readings for the Behavioral Sciences, New York-London-Toronto 1971.
- Suppes, P., Probabilistic Grammars for Natural Languages, Institute for Mathematical Studies in the Social Sciences, Stanford University, California, Psychology Series, Technical Report No. 157, 1970 (vervielfältigtes Manuskript).
- Ungeheuer, G., Language in the light of the information theory; in: International Social Science Journal, Vol. XIX, N. 1.
- Wagenführ, R., Statistik leicht gemacht. Einführung in die deskriptive Statistik, 4. Aufl., Köln 1963.
- Wallis, W.A. and H.V. Roberts, Methoden der Statistik. Ein neuer Weg zu ihrem Verständnis, Freiburg i.Br. o.J. (1959).
- Whittaker, E., From Euclid to Eddington, Cambridge 1949.
- Wiener, N., Kybernetik, Düsseldorf und Wien 1963.
- Wright Mills, C., Kritik der soziologischen Denkweise, Soziologische Texte, Bd. 8, Neuwied/Rhein und Berlin-Spandau 1963.
- Yamane, T., Statistics: An Introductory Analysis, 2nd ed., New York-Evanston-London-Tokyo 1964.
- Yule, G.U., The Statistical Study of Literary Vocabulary, Cambridge 1964.

